

Leonie Felica Tatzel

Verbesserungen beim Laserschneiden mit Methoden des maschinellen Lernens

Leonie Felica Tatzel

**Verbesserungen beim Laserschneiden
mit Methoden des maschinellen Lernens**

Forschungsberichte aus der Industriellen Informationstechnik
Band 24

Institut für Industrielle Informationstechnik
Karlsruher Institut für Technologie
Hrsg. Prof. Dr.-Ing. Michael Heizmann

Eine Übersicht aller bisher in dieser Schriftenreihe erschienenen Bände
finden Sie am Ende des Buchs.

Verbesserungen beim Laserschneiden mit Methoden des maschinellen Lernens

von
Leonie Felica Tatzel

Karlsruher Institut für Technologie
Institut für Industrielle Informationstechnik

Verbesserungen beim Laserschneiden
mit Methoden des maschinellen Lernens

Zur Erlangung des akademischen Grades einer Doktor-Ingenieurin
von der KIT-Fakultät für Elektrotechnik und Informationstechnik des
Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Leonie Felica Tatzel

Tag der mündlichen Prüfung: 10. Juni 2021
Hauptreferent: Prof. Dr.-Ing. M. Heizmann, KIT
Korreferent: Prof. Dr.-Ing. M. Zäh, Technische Universität München

Impressum



Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.
Reprint using the book cover is not allowed.

www.ksp.kit.edu



*This document – excluding parts marked otherwise, the cover, pictures and graphs –
is licensed under a Creative Commons Attribution-Share Alike 4.0 International License
(CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>*



*The cover page is licensed under a Creative Commons
Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0):
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2022 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 2190-6629
ISBN 978-3-7315-1128-1
DOI 10.5445/KSP/1000137690

Vorwort

Die vorliegende Arbeit entstand während meiner dreijährigen Tätigkeit bei der TRUMPF Werkzeugmaschinen GmbH + Co. KG in Ditzingen. Betreut wurde sie von Prof. Dr.-Ing. Michael Heizmann am Institut für Industrielle Informationstechnik (IIIT) des Karlsruher Instituts für Technologie. Ohne die Unterstützung zahlreicher Personen hätte die Arbeit in dieser Form nicht realisiert werden können. Ihnen möchte ich meinen Dank aussprechen.

Zuallererst möchte ich Prof. Dr.-Ing. Michael Heizmann herzlich danken für die Übernahme der Betreuung nach dem überraschenden Tod von Prof. Puento León im Juli 2020. Seine motivierenden Worte und die konstruktive Kritik waren mir in den letzten Monaten eine große Hilfe. Mein Dank gilt auch Prof. Dr.-Ing. Michael Zäh für die Übernahme des Korreferats und die ausführlichen Anmerkungen zur schriftlichen Ausarbeitung.

Dr. Thomas Schneider danke ich für das große Interesse an meiner Arbeit und seinen unermüdlichen Optimismus. Für die Schaffung der wohl besten Rahmenbedingungen, die man sich für eine Industriepromotion wünschen kann, sowie für sein großes Vertrauen in meine Forschung danke ich Dr. Jens Ottnad.

Allen Mitarbeitenden des IIIT danke ich für den offenen Austausch bei den Sommerseminaren und für das Korrekturlesen meiner Arbeit. Bedanken möchte ich mich auch bei allen Studierenden, die ich betreuen durfte und die zu den Ergebnissen dieser Arbeit beigetragen haben. Besonders hervorzuheben sind hier Tobias Haueise und Omar Al Tamimi.

Bei TRUMPF danke ich allen Kolleginnen und Kollegen für die gute Arbeitsatmosphäre und dass sie mir jederzeit mit Rat und Tat zur Seite standen. Großer Dank gebührt Frederick Struckmeier für die guten Diskussionen und den immerwährenden Beistand.

Den größten Dank möchte ich an meine Eltern, Lukas sowie an Ann-Katrin und Felix richten. Danke für eure vielfältige, großartige Unterstützung!

Zusammenfassung

Die fertigende Industrie wird mit einer komplexen Wettbewerbssituation, globalisierten Märkten und kurzen Innovationszyklen konfrontiert. Gleichzeitig ist sie einem großen Kostendruck und teurer werdenden Ressourcen ausgesetzt. Diesen Herausforderungen kann sie begegnen, indem sie erstens auf individuelle Kundenanforderungen eingeht, zweitens Ressourcen sparsam einsetzt und drittens bereits kleine Losgrößen zu günstigen Preisen produziert. Laserschneiden von Metallen ist ein weit verbreiteter Fertigungsprozess. Obwohl er aufgrund seiner hohen Flexibilität dazu prädestiniert ist, diese Anforderungen zu erfüllen, gibt es hier noch immer ein erhebliches Verbesserungspotential. Dieses wird in der vorliegenden Arbeit identifiziert und anschließend mit Methoden des maschinellen Lernens (ML) adressiert.

Als Grundlage wurde eine breite Datenbasis aus fast 4000 Schnittkanten durch die Variation vier dominanter Prozessparameter generiert. Zu jeder Schnittkante sind die verwendeten Prozessparameter, die zugehörigen RGB-Bilder und 3D-Punktwolken sowie die daraus abgeleiteten objektiven Qualitätsmerkmale vorhanden. Da die 3D-Punktwolken fehlerbehaftet sind, wurden verschiedene Methoden zur Verbesserung der Datenqualität erprobt.

Damit individuelle Anforderungen an die Schnittqualität formuliert und überprüft werden können, ist eine objektive und überprüfbare Qualitätsdefinition nötig. Die nach Norm vorgeschriebenen Größen können jedoch nur mit einem Oberflächenmessgerät bestimmt werden. Da ein solches Messgerät in der Produktion meist nicht vorhanden ist, wird die Qualität bisher subjektiv durch den Menschen beurteilt. Inwieweit eine Einzelperson in der Lage ist, diese Beurteilung durchzuführen und welche Merkmale der Schnittkante für den Menschen besonders relevant sind, wurde mithilfe einer Expertenbefragung evaluiert. Am Beispiel der Rautiefe wurde gezeigt, dass die Normgrößen anstatt mit einem teuren Messgerät auch mit einem faltenden neuronalen Netz (engl. *convolutional*

neural network, CNN) aus RGB-Bildern abgeleitet und so im Produktionsalltag zur Verfügung gestellt werden könnten.

Ressourcen- und kostensparend zu produzieren bedeutet unter anderem, Ausschuss zu vermeiden und Maschinenstillstandszeiten sowie die Notwendigkeit menschlicher Eingriffe zu minimieren. Die bisherige manuelle Überprüfung des Zustands einer Laserschneidmaschine entspricht diesen Anforderungen aber keineswegs. Deshalb wurde ein Konzept entwickelt, mit dem der Maschinenzustand zuverlässig und automatisiert überprüft werden kann. Die entscheidende Information liefern die Ist-Prozessparameter. Diese werden mithilfe eines eigens entwickelten CNNs aus dem RGB-Bild der Schnittkante abgeleitet. Aus der Differenz zwischen den Ist- und den eingestellten Soll-Parametern kann auf Probleme in der Maschine geschlossen werden. Da der Mensch nicht in der Lage ist, die Ist-Prozessparameter zu identifizieren, zeigt dieses Beispiel eindrucksvoll das Potential datenbasierter Methoden im Laserschneiden.

Kleine Stückzahlen günstig herzustellen ist nur möglich, wenn man den Gesamtprozess mit allen Einflussgrößen vollständig beherrscht. Aufgrund der enormen Komplexität des Laserschneidprozesses sollte der Mensch durch Modellierungswerkzeuge unterstützt werden. Trotz der jahrzehntelangen Forschung finden diese (meist analytischen oder numerischen Modelle) nahezu keine Anwendung in der Praxis. Stattdessen ist der Mensch, z.B. bei der Ermittlung geeigneter Prozessparameter, auf sich allein gestellt. In dieser Arbeit wurde der Zusammenhang zwischen den variierten Prozessparametern und drei relevanten Größen, Fehlschnittisiko, Rautiefe und Grathöhe, analysiert und modelliert. Des Weiteren wurden CNNs, welche die Zusammenhänge zwischen einem Bild der Schnittkante und jeweils einem der Prozessparameter erlernt hatten, visualisiert. Die für die Entscheidungen des jeweiligen CNNs besonders relevanten Merkmale des Eingangsbildes wurden mithilfe der sogenannten Layer-wise Relevance Propagation (LRP) sichtbar gemacht. Die Visualisierung kann die Akzeptanz des Einsatzes künstlicher neuronaler Netze verbessern und gleichzeitig dazu beitragen, das Prozessverständnis zu erweitern.

Inhaltsverzeichnis

Vorwort	i
Zusammenfassung	iii
Symbolverzeichnis	ix
1 Einleitung	1
1.1 Problemstellung und eigener Beitrag	4
1.1.1 Qualitätsbeurteilung	4
1.1.2 Überprüfung des Maschinenzustands	6
1.1.3 Erweiterung des Prozessverständnisses	8
1.2 Struktur der Arbeit	10
2 Laserschneiden von Metallen	11
2.1 Grundlagen des Laserschneidens	11
2.1.1 Laser in der Materialbearbeitung	12
2.1.2 Laserschneidprozess	12
2.1.3 Laserschneidmaschine	16
2.2 Qualität thermischer Schnitte	18
2.2.1 Schnittkantenqualität nach Norm	18
2.2.2 Nachbearbeitung	21
2.3 Modellierung des Laserschneidprozesses	22
2.3.1 Physikalisch motivierte Modellierung	23
2.3.2 Datenbasierte Modellierung	25
2.4 Laserschneiden in der Praxis	26
2.4.1 Beurteilung der Qualität	27
2.4.2 Überprüfung des Maschinenzustands	27
2.4.3 Ermittlung und Optimierung der Prozessparameter	28
2.5 Diskussion des Stands der Technik	30

3	Methoden des maschinellen Lernens	33
3.1	Grundlagen	33
3.1.1	Wichtige Begriffe	34
3.1.2	Evaluierung der Modellgüte	39
3.1.3	Arbeitsablauf eines ML-Projektes	43
3.2	Künstliche neuronale Netze	44
3.2.1	Grundlagen künstlicher neuronaler Netze	45
3.2.2	Faltende neuronale Netze	48
3.2.3	Implementierungsdetails	51
3.2.4	Highway-Schicht	54
3.2.5	Anwendungen neuronaler Netze in der Fertigung	55
3.2.6	Interpretierbarkeit neuronaler Netze	58
4	Generierung der Datenbasis	63
4.1	Schneidexperiment	63
4.1.1	Probe	66
4.1.2	Material und Schachtelung	67
4.1.3	Überwachung des Maschinenzustands	68
4.2	Objektive Merkmale der Schnittkante	69
4.2.1	Messmethodik	69
4.2.2	RGB-Bild und 3D-Punktwolke	74
4.2.3	Objektive Merkmale der Schnittfläche	76
4.2.4	Objektive Merkmale des Grats	78
4.3	Ausreißerkorrektur	80
4.3.1	Korrektur mit einem Autoencoder	81
4.4	Konzept der Expertenbefragung	85
4.5	Übersicht über die Datenbasis	86
5	Modellierung und Visualisierung	89
5.1	Bildbasierte Modellierung mit neuronalen Netzen	89
5.1.1	Regression der Prozessparameter	89
5.1.2	Klassifikation nach den Faktorstufen eines Prozess- parameters	93
5.1.3	Schätzung der Rautiefe	94
5.1.4	Konfigurationen	95
5.2	Visualisierung mit Layer-wise Relevance Propagation	96
5.2.1	Deep Taylor Decomposition	97

5.2.2	Implementierungsdetails	102
5.3	Zusammenhang zwischen Parametern und Kantenqualität	104
6	Ergebnisse und Diskussion	107
6.1	Datenbasis	107
6.1.1	Fehlschnitte	107
6.1.2	Rautiefe	109
6.1.3	Rechtwinkligkeitstoleranz	119
6.1.4	Grathöhe	120
6.2	Qualitätsbeurteilung	125
6.2.1	Expertenbefragung	125
6.2.2	Bildbasierte Schätzung der Rautiefe	135
6.3	Überprüfung des Maschinenzustands	141
6.3.1	Bildbasierte Regression der Prozessparameter	142
6.4	Erweiterung des Prozessverständnisses	148
6.4.1	Zusammenhänge zwischen Schnittkante und Prozessparametern	149
6.4.2	Analyse und Modellierung des Fehlschnitttrisikos	156
6.4.3	Analyse und Modellierung der Rautiefe	161
6.4.4	Analyse und Modellierung des Grats	164
6.4.5	Praktische Anwendung der Modelle	169
7	Schluss	171
7.1	Ausblick	174
A	Anhang	177
A.1	Technische Daten	177
A.2	LTT-Teile	179
A.3	Schnittkanten der Teile 0949, 0992, 1006, 1056, 1077	180
A.4	LRP-Regeln	181
A.4.1	Kombination der Regeln	182
A.5	Analyse weiterer Einflussgrößen auf die Grathöhe	183
A.6	Modellierung der Rautiefe	186
	Literaturverzeichnis	189
	Eigene Veröffentlichungen	204
	Betreute studentische Arbeiten	205

Symbolverzeichnis

Allgemeine Abkürzungen

Abkürzung Bedeutung

2D	Zweidimensional
3D	Dreidimensional
ADB	Abstand Düse-Blech
AE	Autoencoder
BCA	Gewichtete Korrektklassifikationsrate, engl. <i>balanced classification accuracy</i>
BN	Batch-Normalisierung
bzw.	Beziehungsweise
CA	Korrektklassifikationsrate, engl. <i>classification accuracy</i>
ca.	Circa
CCE	Kategorische Kreuzentropie, engl. <i>categorical cross entropy</i>
CMOS	Halbleiterbauelement mit n- und p-Transistor auf einem Substrat, engl. <i>complementary metal-oxide-semiconductor</i>
CNN	Faltendes neuronales Netz, engl. <i>convolutional neural network</i>
d. h.	Das heißt
DMC	Data Matrix Code
DTD	Taylorreihe für neuronale Netze, engl. <i>deep taylor decomposition</i>
dt.	Deutsch
E1 bis E5	Experte 1 bis Experte 5
EM	Einstellmaß
engl.	Englisch
EU	Europäische Union
FC	Vollständig verbundene Schichten, engl. <i>fully connected layers</i>
FP	Fokusposition

Abkürzung Bedeutung

FS	Fehlschnitt
GAP	Global Average Pooling
ges.	Gesamt
ggf.	Gegebenenfalls
ILT	Fraunhofer-Institut für Lasertechnik
KI	Künstliche Intelligenz
KNN	Künstliches neuronales Netz
LED	Leuchtdiode, engl. <i>light emitting diode</i>
LRP	Layer-wise Relevance Propagation
LTT	Laser-Technologie-Tabelle
MAE	Mittlerer absoluter Fehler, engl. <i>mean absolute error</i>
Mio.	Million
ML	Maschinelles Lernen, engl. <i>machine learning</i>
Mrd.	Milliarde
MSE	Mittlerer quadratischer Fehler, engl. <i>mean squared error</i>
MW	Mittelwert
PReLU	Engl. <i>parameterized rectified linear unit</i>
ReLU	Engl. <i>rectified linear unit</i>
RGBD	Rot-Grün-Blau-Farbraum und Höheninformation
RGB	Rot-Grün-Blau-Farbraum
SGD	Stochastisches Gradientenverfahren, engl. <i>stochastic gradient descent</i>
Std	Standardabweichung
TRUMPF	TRUMPF Werkzeugmaschinen GmbH + Co. KG
US	Vereinigte Staaten von Amerika, engl. <i>united states</i>
z. B.	Zum Beispiel

Symbole

Lateinische Buchstaben

Symbol	Bedeutung
$a, \mathbf{a}, \mathbf{A}$	Aktivierung eines Neurons, Vektor/Matrix der Aktivierungen der Neuronen einer Schicht
$b, \mathbf{b}, \mathbf{B}$	Biasterm eines Neurons, Vektor/Matrix der Biasterme aller Neuronen einer Schicht
B	Block, Zusammenfassung mehrerer Schichten eines KNN
c	Spezifische Wärmekapazität
d	Dicke des Werkstücks bzw. des Blechs
E	Fehler-/Verlust-/Kostenfunktion
h	Höhendifferenz
h_S	Spezifische Schmelzenthalpie
\mathbf{K}	Faltungskern
L	Anzahl der Schichten eines KNN
M	Anzahl abhängiger Variablen (Output)
N	Anzahl Datenpunkte
P	Anzahl unabhängiger Variablen (Input)
p	Druck des Schneidgases
P_V	Verlustleistung
P_a	Absorbierte Leistung
P_{ch}	Chemische Verbrennungsleistung
R	Relevanz eines Neurons
\hat{R}	Skalierte Relevanz eines Neurons
Rz	Gemittelte Rautiefe
s	Schnittspaltbreite
T	Temperatur
u	Rechtwinkligkeitstoleranz
v	Vorschub, Schneidgeschwindigkeit
$w, \mathbf{w}, \mathbf{W}$	Modellparameter, Vektor/Matrix der Modellparameter
$x, \mathbf{x}, \mathbf{X}$	Unabhängige Variable(n), Input
y, \mathbf{y}	Abhängige Variable(n), Output
$z, \mathbf{z}, \mathbf{Z}$	Wert eines Neurons, Vektor/Matrix der Werte der Neuronen einer Schicht vor Anwendung der Aktivierungsfunktion

Griechische Buchstaben

Symbol	Bedeutung
χ	Neue Standardabweichung nach Batch-Normalisierung
δ	Verschiebung, Differenz
ϵ	Stabilisierungssummand in der LRP- ϵ -Regel
η	Terme 2. und höherer Ordnung bei der Taylor-Reihenentwicklung
γ	Gewichtungsfaktor in der LRP- γ -Regel
λ	Lernrate
μ	Erwartungswert
ϕ	Neuer Mittelwert nach Batch-Normalisierung
ψ	Skalierungsfaktor für die Visualisierung der LRP-Heatmaps
ρ	Dichte
σ	Standardabweichung
τ	Zufälliger Fehlerterm
θ	Triangulationswinkel
v	Suchrichtung nach der Entwicklungsstelle

(Hochgestellte) Indizes

Index	Bedeutung
$(\bullet)^{(M_B)}$	Messbereich
$(\bullet)^{(l)}$	Schicht eines KNNs
$(\bullet)^{(l,l+1)}$	Übergang zwischen der Schicht l und der Schicht $l + 1$ eines KNNs

Indizes

Index	Bedeutung
$(\bullet)_i$	i -tes Neuron (der Schicht l)
$(\bullet)_{ij}$	Verbindung des i -ten mit dem j -ten Neuron

Index	Bedeutung
$(\bullet)_j$	j -tes Neuron oder j -ter Biasterm (der Schicht $l + 1$)
$(\bullet)_m$	m -te abhängige Variable (Output)
$(\bullet)_{M_L}$	Messlinie bei der Rautiefenmessung
$(\bullet)_n$	n -ter Eingangsdatenpunkt, n -tes Trainingsbeispiel
$(\bullet)_p$	p -te unabhängige Variable, z. B. p -tes Pixel des Eingangsbildes
$(\bullet)_{l,l+1}$	Übergang zwischen der Schicht l und der Schicht $l + 1$ eines KNNs
$(\bullet)_u$	u -te Feature Map (der Schicht l)
$(\bullet)_{uv}$	Verbindung der u -ten Feature Map (der Schicht l) mit der v -ten Feature Map (der Schicht $l + 1$)
$(\bullet)_v$	v -te Feature Map (der Schicht $l + 1$)

Mathematische Operatoren

Operator	Bedeutung
*	Faltung
o	Hadamard Produkt

1 Einleitung

Die Idee, Materialien mit Licht zu bearbeiten, hat Menschen seit dem Tag fasziniert, als sie das erste Mal Papier mit Hilfe des Sonnenlichts und einer Lupe verbrannten. Heutzutage wird zur Materialbearbeitung oft Laserstrahlung eingesetzt.

Das Prinzip der Erzeugung von Laserstrahlung wurde bereits 1917 durch Albert Einstein postuliert [Ein17] und ungefähr 40 Jahre später durch Theodore Maiman mit einem ersten Prototypen des Lasers in die Tat umgesetzt [Mai60]. Damit hat eine Erfolgsgeschichte begonnen, die bis heute nicht abgeschlossen ist. Neben der Erforschung, Weiterentwicklung und Anwendung des Lasers in der Wissenschaft hat er sich in vielen anderen Bereichen etabliert. Heute sind Laser aus der Industrie (z. B. Laserschweißen von Karosserieteilen), der Medizin (z. B. refraktive Chirurgie), der Informationstechnologie (z. B. Glasfaser-Internet) und aus unserem Alltag (z. B. Laserdrucker) nicht mehr wegzudenken.

Ungefähr ein Drittel aller weltweit verkauften Laser werden in der Materialbearbeitung eingesetzt. 2015 wurde mit Lasern für die Materialbearbeitung ein Umsatz von 3,8 Mrd. US-Dollar erzielt, 2018 waren es bereits 5,6 Mrd. US-Dollar. Den davon mit Abstand größten Anteil von 35 % machen Laserschneidprozesse aus. Weitere häufige Anwendungen sind Schweißen (16 %), Markieren (15 %) und die Herstellung von Halbleitern und Displays (14 %). [Kin+18]

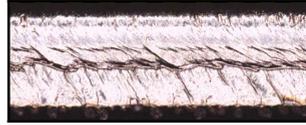
Beim Laserschneiden von Metallen wird das Werkstück durch einen fokussierten Laserstrahl erhitzt und geschmolzen oder verdampft. Gleichzeitig wird das geschmolzene Material mithilfe des sogenannten Schneidgases entfernt. Durch die Relativbewegung zwischen Laserstrahl und Werkstück entsteht ein Schnittspalt. Die resultierende Schnittkante hängt von den Prozessparametern, dem Zustand der Laserschneidmaschine und den Materialeigenschaften des Werkstücks ab [SM10]. Zwei exemplarische Schnittkanten sind in Abbildung 1.1 vor einem schwarzen Hintergrund zu sehen. In beiden Fällen handelt es sich um ein 3 mm dickes Edelstahlblech.

— 1 mm



(a) Glatte Schnittkante ohne Grat.

— 1 mm



(b) Raue Schnittkante mit dickem Grat.

Abbildung 1.1 Zwei unterschiedliche Laserschnittkanten.

Aufgrund unterschiedlich eingestellter Prozessparameter ist die in Abbildung 1.1(a) gezeigte Schnittkante glatt, die in Abbildung 1.1(b) gezeigte Kante weist dagegen eine raue Oberfläche und einen sogenannten Grat an der Unterseite auf.

Beim Laserschneiden handelt es sich um ein sehr etabliertes Fertigungsverfahren, das seit Jahrzehnten industriell eingesetzt wird. Umso überraschender ist es, dass dieser Prozess noch nicht vollständig verstanden ist [HG09; PSW12; SM10]. Bis heute wird an der Modellierung und Optimierung des Prozesses geforscht. Bei den eingesetzten Modellen handelt es sich nahezu ausschließlich um physikalisch motivierte Modelle, die mit analytischen und klassischen numerischen Methoden entwickelt wurden. Aufgrund der enormen Komplexität des Schneidprozesses, die sich aus dem Zusammenspiel verschiedener physikalischer Prozesse und der Vielzahl an Einflussgrößen ergibt (siehe Abschnitt 2.1), modellieren sie nur Teilaspekte und basieren auf vielen vereinfachenden Annahmen. Diese Betrachtung ist der Erweiterung des Prozessverständnisses dienlich, für die Praxis jedoch wenig hilfreich. Methoden des maschinellen Lernens finden kaum Anwendung, obwohl sie geradezu prädestiniert für die Modellierung komplexer Zusammenhänge sind. Angesichts fehlender Alternativen verlässt man sich in der Praxis weitestgehend auf den Menschen:

- Die Qualität einer Schnittkante wird durch Anfassen und Anschauen und nicht basierend auf gemessenen Kriterien beurteilt.
- Der Zustand einer Laserschneidmaschine, welcher große Auswirkungen auf den Schneidprozess haben kann, wird manuell durch Prüfen verschiedener Einzelkomponenten kontrolliert. Dies führt zu langen Maschinenstillstandszeiten.

- Um neue Prozessparameter (z. B. für eine höhere Laserleistung) zu ermitteln oder bestehende zu optimieren, werden aufwendige experimentelle Untersuchungen durchgeführt, die viel Vorwissen und Erfahrung erfordern.

Dieses Vorgehen wird vor dem Hintergrund der veränderten Anforderungen an die fertigende Industrie immer problematischer. Diese wird mit einer komplexen Wettbewerbssituation, globalisierten Märkten und kurzen Innovationszyklen konfrontiert [LRZ06]. Gleichzeitig ist sie einem großen Kostendruck und teurer werdenden Ressourcen ausgesetzt, wie Brecher und Özdemir [BÖ15] schreiben. Die Autoren schlagen vor, diesen Herausforderungen mit

- dem Eingehen auf individuelle Anforderungen,
- einer sparsamen Ressourcennutzung und
- dem Fertigen kleiner Losgrößen zu günstigen Preisen

zu begegnen. Obwohl das Laserschneiden wegen seiner Flexibilität geradezu prädestiniert für die Erfüllung dieser Anforderungen ist, können sie mit der aktuellen Vorgehensweise nicht bewältigt werden. Solange es keine objektive, überall anwendbare Definition der Schnittqualität gibt, können kundenspezifische Anforderungen nicht berücksichtigt werden. Ressourcensparend zu produzieren bedeutet unter anderem, Ausschuss und Maschinenstillstandszeiten zu vermeiden und menschliche Eingriffe zu minimieren. Die manuelle Überprüfung des Zustands einer Laserschneidmaschine entspricht dieser Anforderung keineswegs. Kundenspezifische Produkte in kleinen Stückzahlen und gleichzeitig zu günstigen Preisen zu produzieren, ist nur möglich, wenn man den Gesamtprozess vollständig beherrscht. Aufgrund der großen Komplexität sollte der Mensch dabei von Modellierungswerkzeugen unterstützt werden.

In dieser Arbeit werden die obenstehenden Probleme mithilfe von Methoden des maschinellen Lernens adressiert. Mit dieser Methodik wurden in den letzten Jahren große Erfolge in verschiedensten Anwendungsbereichen erzielt. Für die Modellierung des Laserschneidprozesses wird sie jedoch noch kaum eingesetzt.

1.1 Problemstellung und eigener Beitrag

Um Methoden des maschinellen Lernens einsetzen zu können, benötigt man eine Datenbasis. Da es für das Laserschneiden keine öffentlich verfügbaren Daten gibt, wurde im Rahmen dieser Arbeit eine Datengrundlage geschaffen.

Eigener Beitrag Es wurde eine große Datenbasis bestehend aus 3596 Laserschnittkanten mit ganz unterschiedlichen Schnittbildern generiert (siehe Kapitel 4). Von jeder Schnittkante wurden RGB-Bilder und 3D-Punktwolken aufgenommen. Aus letzteren konnten verschiedene objektive Qualitätsmerkmale berechnet werden. Durch die Verwendung einer modernen Laserschneidmaschine mit einem Festkörperlaser können die Herausforderungen (z. B. durch Prozessstreuungen) in der Praxis realistisch eingeschätzt werden. Neben einer Analyse der Datenqualität wurden Methoden zur Verbesserung derselben evaluiert (siehe Abschnitt 6.1). Die Größe der Datenbasis ermöglicht das Training tiefer neuronaler Netze mit Bildern als Eingangsdaten, was völlig neue Anwendungsmöglichkeiten eröffnet. Damit kann erstmalig das große Potential datenbasierter Methoden im Laserschneiden gezeigt werden.

Im Folgenden wird für jede der drei oben genannten Anforderungen an die fertigende Industrie aufgezeigt, welche Probleme aus dem aktuellen Stand der Technik und dem Vorgehen in der Praxis resultieren. Anschließend wird der eigene Beitrag zur Adressierung dieser Probleme dargestellt. Das Gesamtkonzept der Arbeit ist in Abbildung 1.2 zusammengefasst.

1.1.1 Qualitätsbeurteilung

Zur Formulierung und Überprüfung individueller, kundenspezifischer Anforderungen wird eine objektive Qualitätsdefinition benötigt.

Stand der Technik Die Norm DIN EN ISO 9013 [Deu17] definiert zwei obligatorische Qualitätskriterien für Laserschnittkanten: gemittelte Rautiefe $Rz5$ und Rechtwinkligkeitstoleranz u . In der Literatur wird die Rautiefe

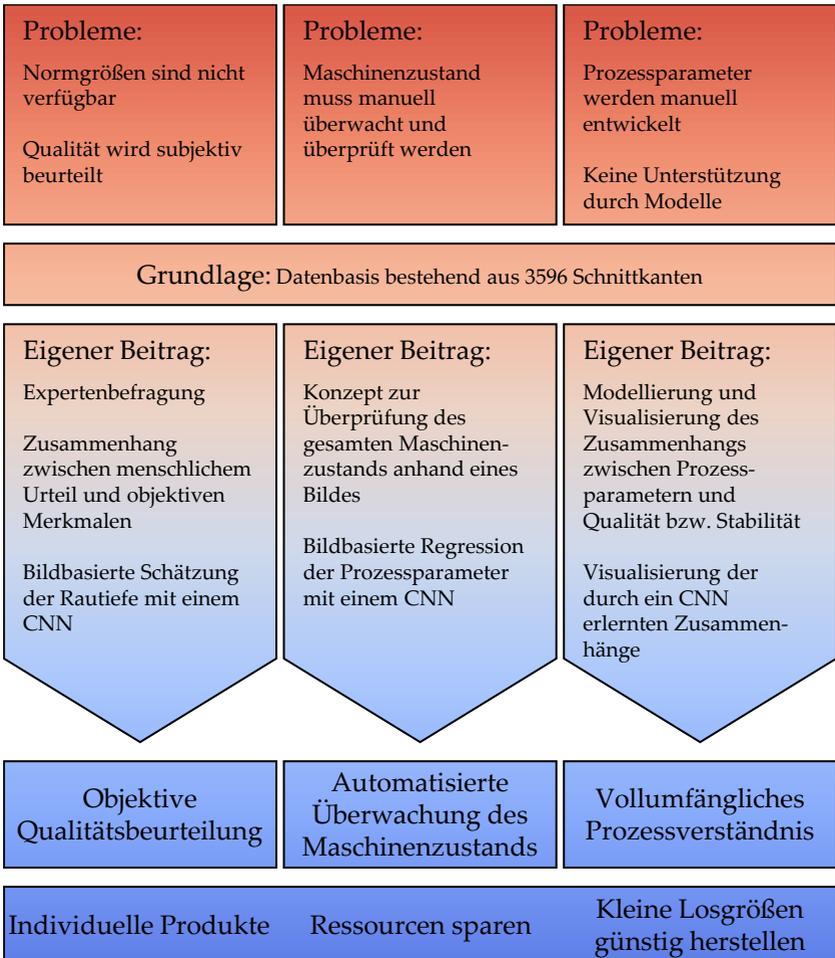


Abbildung 1.2 Gesamtkonzept der Arbeit. Das Ziel sind Verbesserungen im Laserschneiden hinsichtlich der drei Anforderungen: individuelle Produkte, Ressourcen sparen und kleine Losgrößen günstig herstellen.

häufig mit der Schnittkantenqualität gleichgesetzt (z. B. [Ter+16]). Für die Ermittlung der Normgrößen ist ein Oberflächenmessgerät nötig.

Vorgehen in der Praxis In der Praxis ist meist kein Messgerät zur Bestimmung der Normgrößen vorhanden. Die Qualität einer Schnittkante wird vom Menschen durch Anschauen und Anfassen beurteilt. Mit jeder Laserschneidmaschine werden Parametertabellen ausgeliefert, in denen geeignete Prozessparameterkombinationen für häufig verwendete Materialien und Blechdicken zu finden sind.

Problemstellung Da objektive Qualitätsmerkmale in der Praxis nicht ermittelt werden können, wird die Schnittkantenqualität subjektiv von Einzelpersonen bewertet. Diese Qualitätsbeurteilung ist äußerst intransparent, der Zusammenhang zwischen der menschlichen Beurteilung und den Normgrößen ist unklar. Individuelle Anforderungen und Präferenzen können so weder formuliert noch überprüft werden.

Für die Bereitstellung der Parametertabellen wird ein allgemein gültiges Qualitätsverständnis angenommen – über verschiedene Branchen, Produkte und Märkte hinweg. Parameteranpassungen aufgrund abweichenden Qualitätsvorstellungen können nur mit einem hohen manuellen Aufwand und Expertenkenntnissen vorgenommen werden.

Eigener Beitrag Anhand einer Expertenbefragung werden die Probleme des bisherigen Vorgehens in der Praxis aufgezeigt (siehe Abschnitt 6.2.1). Im Rahmen der Befragung wurden 100 Schnittkanten von jeweils fünf Personen bewertet. Damit die subjektive Qualitätsbeurteilung in der Produktion durch eine wichtige objektive Größe ergänzt werden kann, wurde ein CNN entwickelt, mit dem die Rautiefe basierend auf einem RGB-Bild geschätzt werden kann (siehe Abschnitt 6.2.2). Dafür ist kein spezielles Messgerät erforderlich. Eine Alternative zu den Parametertabellen wird in Abschnitt 1.1.3 erwähnt.

1.1.2 Überprüfung des Maschinenzustands

Eine zuverlässige und automatisierte Überprüfung des Maschinenzustands hilft dabei, ressourcenschonend zu produzieren.

Stand der Technik Laserschneidmaschinen können prinzipiell durch automatisierte Belade- und Entladesysteme mit wenigen menschlichen Eingriffen betrieben werden, solange keine Probleme auftreten. Eine Laserschneidmaschine setzt sich aus vielen verschiedenen Einzelkomponenten zusammen, die teilweise mit Sensorik überwacht werden. Ein klar definierter Maschinenzustand ist die Grundvoraussetzung für die Beherrschbarkeit des Laserschneidprozesses und für die Gültigkeit jeglicher Modelle und Prozessparametertabellen, weil etwaige Störgrößen nur dann beherrschbar sind.

Vorgehen in der Praxis Aufgrund der Vielzahl möglicher Ursachen und der Komplexität des Schneidprozesses sind Menschen nicht in der Lage, von einer fehlerhaften Schnittkante auf die Ursache zu schließen. Bei Problemen wird deshalb zunächst der Maschinenzustand überprüft, indem eine Maschinenkomponente nach der anderen manuell kontrolliert wird. Die Sensorik unterstützt dabei, gibt aber nur den Zustand einer Komponente, z. B. den Verschmutzungsgrad, an und nicht die Auswirkungen auf den Gesamtprozess. Da die Überprüfung nicht im laufenden Betrieb erfolgen kann, kommt es zu längeren Maschinenstillstandszeiten. Wenn die Schneidprobleme durch die Wartung der Maschine nicht behoben werden können, müssen in einem zweiten Schritt die Soll-Prozessparameter optimiert werden.

Problemstellung Bislang existiert keine Möglichkeit, den Zustand des Gesamtsystems automatisiert zu überprüfen und ein vorausschauendes Eingreifen ist nicht möglich. Das verringert die Produktivität (d. h. die Teileanzahl pro Zeit) der Maschine und kann Ausschuss oder einen hohen Nachbearbeitungsaufwand verursachen.

Eigener Beitrag Es wurde ein neues Konzept zur Überprüfung des Maschinenzustands entwickelt: Die Ist-Prozessparameter werden mithilfe eines eigens entwickelten CNNs aus dem RGB-Bild der Schnittkante abgeleitet (siehe Abschnitt 6.3.1). Aus der Differenz zwischen den so ermittelten Ist- und den eingestellten Soll-Parametern kann schneller auf Defekte oder Verschmutzungen in der Maschine geschlossen werden. Vorausschauende Wartung könnte durch eine permanente, automatisierte Überwachung

der Maschine und durch die Detektion kleinster Veränderungen realisiert werden.

Es sind keine Arbeiten bekannt, in denen von der Schnittkante auf die zugrunde liegenden Parameter geschlossen wird. Da der Mensch zu diesem Rückschluss aufgrund der Komplexität des Prozesses nicht in der Lage ist, glauben viele, dass es gar keinen eindeutigen Zusammenhang gibt. Während oft versucht wird, mittels künstlicher Intelligenz (KI) an die Leistungsfähigkeit des Menschen heranzukommen (z. B. bei der Gesichtserkennung), wurde hier ein Zusammenhang modelliert, der für den Menschen unverständlich ist.

1.1.3 Erweiterung des Prozessverständnisses

Um ab dem ersten Teil zu günstigen Preisen zu produzieren, ist die vollständige Beherrschung des Produktionsprozesses unabdingbar.

Stand der Forschung – Laserschneiden Neben den oben erwähnten physikalisch motivierten Modellansätzen werden seit einigen Jahren auch Methoden des maschinellen Lernens eingesetzt. Allerdings handelt es sich dabei meist um einfache Modelle wie Entscheidungs bäume oder neuronale Netze mit einer lediglich einstelligen Anzahl verdeckter Neuronen. Diese werden, analog zu den analytischen oder numerischen Ansätzen, eingesetzt, um die Zusammenhänge zwischen verschiedenen Prozessparametern und einzelnen Qualitätskriterien zu modellieren. Als Trainingsdaten werden entweder Simulationsdaten oder experimentell generierte Datensätze mit weniger als 50 Beispielen verwendet.

Vorgehen in der Praxis Die Ermittlung neuer und die Anpassung bestehender Prozessparameter erfolgen manuell und erfordern viel Erfahrung und Zeit. Deshalb werden die bereits erwähnten Parametertabellen angeboten.

Stand der Technik – Maschinelles Lernen Das Thema künstliche Intelligenz ist medial sehr präsent und Methoden des maschinellen Lernens, insbesondere künstliche neuronale Netze, erzielen große Erfolge in verschiedensten Bereichen wie z. B. dem autonomen Fahren [AI-+17] und

der Medizin [Est+17]. Im Produktionskontext werden faltende neuronale Netze beispielsweise zur Erkennung und Lokalisierung von Defekten [Lin+19], zur vorausschauenden Wartung [KC19] und zur Ermittlung von Oberflächenkenngrößen [Sun+18] eingesetzt. Als Eingangsdaten werden meist zeitabhängige Sensorsignale oder Bilder verwendet. Eine grundsätzliche Herausforderung tiefer neuronaler Netze ist die Interpretierbarkeit der erlernten Zusammenhänge.

Problemstellung Die enormen Fortschritte, die in vielen Bereichen mit Methoden des maschinellen Lernens erzielt wurden, kommen beim Laserschneiden nicht an. Datenbasierte Methoden entfalten ihr Potential vor allem bei der Verwendung komplexer, nicht linearer Modelle (z. B. tiefer neuronaler Netze) in Kombination mit einer breiten Datenbasis.

Da bestehende Modelle praxisrelevante Fragestellungen nur unzureichend beantworten, verlässt man sich vollständig auf die Parametertabellen und auf erfahrene Fachkräfte. Die Tabellen führen jedoch oft zu nicht zufriedenstellenden Ergebnissen. Wegen des manuellen Vorgehens bleibt der Aufwand für die Ermittlung neuer Parameter immer derselbe.

Eigener Beitrag Die Erweiterung des Prozessverständnisses und die Unterstützung des Menschen durch die Bereitstellung nützlicher Informationen und geeigneter Modellierungswerkzeuge sind die zentralen Themen dieser Arbeit und finden sich in vielen Teilaspekten wieder.

Am Beispiel von CNNs, die den Zusammenhang zwischen einem Bild der Schnittkante und jeweils einem der Prozessparameter erlernt haben, wird gezeigt, dass sich die Verwendung datenbasierter Modelle und die Erweiterung des Verständnisses nicht ausschließen. Eine geeignete Visualisierungstechnik kann die Akzeptanz künstlicher neuronaler Netze stärken, das Prozessverständnis erweitern und den Menschen bei der Lösung von Schneidproblemen unterstützen (siehe Abschnitt 6.4.1).

Die Zusammenhänge zwischen den Prozessparametern und der Prozessstabilität sowie mehreren Qualitätsmerkmalen der Schnittkante werden analysiert und modelliert (siehe Abschnitte 6.4.2, 6.4.3, 6.4.4). Diese Modelle könnten die Grundlage für die individuelle Optimierung des Prozesses bilden.

1.2 Struktur der Arbeit

Die vorliegende Arbeit ist wie folgt aufgebaut. Kapitel 2 erläutert die Grundlagen des Laserschneidens und die objektiven Kriterien zur Beurteilung der Schnittkantenqualität. Es wird auf den Stand der Wissenschaft bezüglich der Modellierung des Laserschneidprozesses eingegangen. Dabei werden physikalisch motivierte und datenbasierte Modellansätze unterschieden. Die in der Praxis üblichen Vorgehensweisen für die Beurteilung der Schnittqualität, die Überprüfung des Maschinenzustands und die Ermittlung neuer bzw. die Optimierung bestehender Prozessparameter werden erklärt. Die aktuellen Herausforderungen und die daraus abgeleiteten notwendigen Veränderungen im Zuge der vierten industriellen Revolution werden thematisiert. Zum Abschluss des Kapitels wird diskutiert, warum datenbasierte Methoden im Laserschneiden bislang kaum eingesetzt werden. Anschließend folgen im Kapitel 3 die Grundbegriffe des maschinellen Lernens, ergänzt durch eine Einführung in faltende neuronale Netze und Autoencoder (AE). Es wird auf den Stand der Technik von ML-Anwendungen im industriellen Umfeld eingegangen und auf den Stand der Forschung zur Interpretierbarkeit neuronaler Netze. Kapitel 4 stellt die Datenbasis vor. Es wird erklärt, mit welchen Prozessparametern die verschiedenen Schnittkanten generiert wurden, wie sie vermessen und die RGB-Bilder aufgenommen wurden. Ein Abschnitt widmet sich verschiedenen Methoden zur Korrektur von Rautiefenmessfehlern. Außerdem wird das Konzept der Expertenbefragung erläutert. In Abschnitt 4.5 befindet sich ein Überblick über die verfügbaren Daten. In Kapitel 5 werden die entwickelten CNN-Architekturen und die Visualisierungstechnik Layer-wise Relevance Propagation vorgestellt. Die Ergebnisse werden in Kapitel 6 präsentiert. Das Kapitel ist entsprechend der Themenblöcke in Abschnitt 1.1 gegliedert. Kapitel 7 fasst die wichtigsten Ergebnisse zusammen und schließt die Arbeit mit einem Ausblick ab.

Im Kontext des maschinellen Lernens sind die geläufigen Begriffe häufig englisch. Es wird versucht, deutsche Übersetzungen zu verwenden, sofern der Bezug zum englischen Fachbegriff klar ersichtlich bleibt. In allen anderen Fällen werden die Originalbegriffe wie deutsche Worte verwendet und bei ihrer ersten Nennung übersetzt bzw. erklärt.

2 Laserschneiden von Metallen

In diesem Kapitel werden zunächst die Grundlagen des Laserschneidens erläutert. Anschließend wird darauf eingegangen, wie die Qualität einer Laserschnittkante beurteilt wird. Der Stand der Technik bei der Modellierung des Schneidprozesses wird aufgezeigt und diskutiert. Ergänzt wird das Kapitel durch eine Schilderung der Vorgehensweise in der Praxis.

2.1 Grundlagen des Laserschneidens

Das Wort Laser bedeutet *Light Amplification by Stimulated Emission of Radiation* (dt. Lichtverstärkung durch stimulierte Emission von Strahlung). Sowohl der physikalische Effekt als auch das Gerät, mit dem Laserstrahlung generiert wird, werden als Laser bezeichnet.

Grundsätzlich besteht jeder Laser aus einem aktiven Lasermedium, einem Pumpmechanismus und einem Laserresonator. Als Medium kann beispielsweise ein Gas oder ein Kristall verwendet werden. Voraussetzung ist ein Material mit mindestens drei atomaren oder molekularen Energieniveaus, damit durch Pumpen mit Energie die sogenannte Besetzungsinversion hergestellt werden kann. In diesem Fall ist der höhere Zustand des optischen Übergangs (Laserniveau) mit einer größeren Wahrscheinlichkeit besetzt als der darunter liegende Grundzustand. Das dritte Energieniveau befindet sich energetisch über den beiden anderen und entleert sich strahlungsfrei auf das Laserniveau. Die Energiedifferenz zwischen Laserniveau und Grundzustand bestimmt die Wellenlänge der entstehenden Strahlung. Das Pumpen kann optisch (z. B. mit einer Lampe) oder elektrisch (z. B. über Gasentladung) erfolgen. Das Lasermedium befindet sich innerhalb eines Laserresonators, welcher mit optischen Elementen für eine resonante Rückkopplung für bestimmte Wellenlängen und Richtungen sorgt. Die Strahlung, die anfänglich durch spontane Emission entsteht, wird so immer wieder durch das Lasermedium geleitet und löst stimulierte

Emission aus. Ein Teil der Strahlung wird auf einer Seite des Resonators ausgekoppelt. Das ist der eigentliche Laserstrahl. [Sve+12]

Laserstrahlen sind elektromagnetische Wellen, die sich durch die Kombination der folgenden Eigenschaften auszeichnen: quasi monochromatisch, große Kohärenzlänge, geringe Divergenz und große Leistungsdichte. Laserstrahlung kann in verschiedenen Bereichen des elektromagnetischen Frequenzspektrums und mit ganz unterschiedlichen Leistungen erzeugt werden. Seit der Konstruktion eines ersten Prototyps 1960 durch Maiman [Mai60] werden Laser in den verschiedensten Bereichen eingesetzt.

2.1.1 Laser in der Materialbearbeitung

Im industriellen Umfeld spielen Laser als Werkzeug eine große Rolle. Es werden verschiedene Lasertypen eingesetzt, die nach der Art des jeweiligen aktiven Mediums benannt sind. Beispiele sind Gas-, Festkörper-, Dioden-, Faser-, Excimer- und Farbstofflaser. Zu den vielfältigen Einsatzmöglichkeiten zählen Schneiden, Bohren, Schweißen, Oberflächenbehandlung, Prototypenbau, abtragende Verfahren und Reinigung [SM10].

Laserschneiden ist die häufigste industrielle Anwendung des Lasers [Kin+18] und hat bestehende Technologien in vielen Bereichen ersetzt. Der Hauptgrund dafür ist, dass beim Trennen mit einem Laser in vielen Fällen eine höhere Produktivität und Qualität erreicht werden kann als mit alternativen Methoden. Powell und Wykes [PW89] vergleichen den Laser als Schneidwerkzeug ausführlich mit herkömmlichen Trennverfahren. Die wichtigsten Vorteile des Laserschneidens sind die große Flexibilität hinsichtlich Formen und Materialien, der schmale Schnittspalt, rechtwinklige Ecken und vergleichsweise saubere Kanten mit einer kleinen Wärmeeinflusszone. Des Weiteren sind mit Lasern hohe Bearbeitungsgeschwindigkeiten möglich und aufgrund des kraftfreien Schneidens ist keine Klemmung des Werkstücks erforderlich. Die folgenden Ausführungen beziehen sich auf das Laserschneiden metallischer Werkstoffe.

2.1.2 Laserschneidprozess

In Abbildung 2.1 ist das Grundprinzip des Laserschneidens veranschaulicht. Mehrere Optiken fokussieren den Laserstrahl im sogenannten Schneidkopf der Laserschneidmaschine. Der Strahl tritt dann zusammen

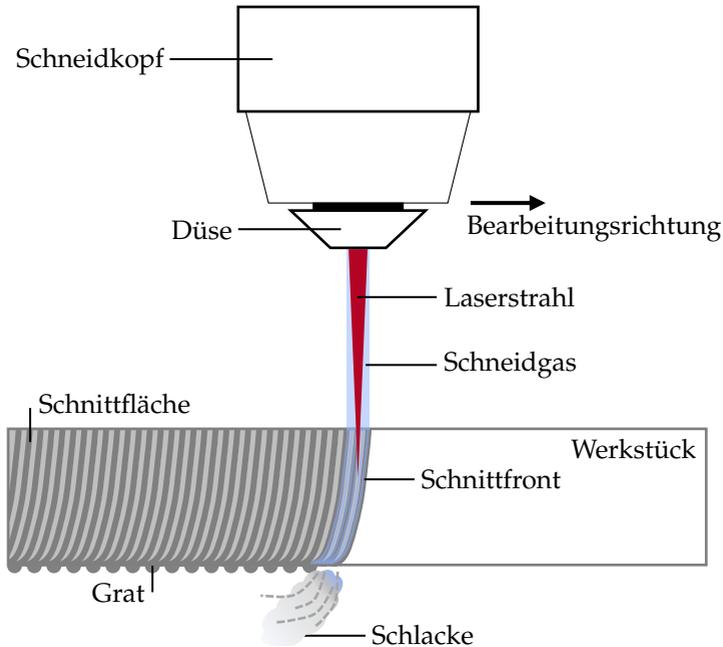


Abbildung 2.1 Prinzip des Laserschneidens, Abbildung nach Hügel und Graf [HG09].

mit einem (üblicherweise) koaxialen Gasstrom aus der Düse des Schneidkopfs aus. Der Strahl trifft auf eine geneigte Schnittfront, an der er das Material (von oben nach unten) schmilzt und teilweise verdampft. Das geschmolzene Material wird, unterstützt durch die Strömung des Schneidgases, entfernt. Aufgrund der Relativbewegung zwischen Laserstrahl und Werkstück entsteht ein Schnittspalt. Die so entstandenen Schnittkanten weisen eine Riefenstruktur auf der Schnittfläche sowie gegebenenfalls Gratperlen an der Unterseite auf.

1967 wurde erstmals gezeigt, dass ein CO_2 -Gaslaser Metall schneiden kann [SH67]. Anschließend wurde die Technologie jahrzehntelang eingesetzt. Mittlerweile werden die Gaslaser aber vermehrt durch Festkörper- oder Faserlaser ersetzt. Diese haben eine kürzere Wellenlänge, lassen sich besser fokussieren und weisen einen höheren Absorptionsgrad in metallischen Werkstoffen auf, was eine erhöhte Schneidgeschwindigkeit und das

Trennen von Buntmetallblechen ermöglicht [HG09; Ols11]. Ein weiterer Vorteil gegenüber CO₂-Lasern ist, dass die Laserstrahlung von Festkörper- und Faserlasern mit einem Lichtleitkabel transportiert werden kann. Der große Nachteil ist die verringerte Qualität der Schnittkante [Ols11]. In dieser Arbeit wurde ein Scheiben-Festkörperlaser eingesetzt (TruDisk 12001). Dessen aktives Medium ist ein Ytterbium dotierter Yttrium-Aluminium-Granat-Kristall. Dieser erzeugt Laserstrahlung mit einer Wellenlänge von 1030 nm. Technische Details zum TruDisk 12001 sind in Anhang A.1 zu finden.

Unterschieden werden Schmelzschneiden und Brennschneiden. Beim Schmelzschneiden wird das geschmolzene Material mit einem Inertgas (z. B. Stickstoff oder Argon) aus dem Schnittpalt ausgetrieben. Beim Brennschneiden wird Sauerstoff als Schneidgas eingesetzt, was eine exotherme Reaktion mit dem geschmolzenen Material auslöst und zusätzliche Energie freisetzt. Dies wird beispielsweise genutzt, um sehr dicke Bleche zu schneiden. Nachteilig ist die dabei auf der Schnittkante entstehende Oxidschicht. Beim Schneiden von Edelstahl wird deshalb gewöhnlich Stickstoff verwendet. [SM10]

Laserschneiden wird seit Jahrzehnten erfolgreich industriell eingesetzt. Die involvierten physikalischen Vorgänge sind jedoch, aufgrund ihrer großen Komplexität und der Tatsache, dass es schwierig ist, den Prozess direkt zu beobachten, bis heute nicht vollständig verstanden [HG09; PSW12].

2.1.2.1 Energiebilanz

Bevor das Schneiden beginnen kann, muss das Werkstück erhitzt und geschmolzen oder verdampft werden, bis sich ein Loch (engl. *keyhole*) bildet. Dieser Prozess wird als Einstechen bezeichnet. Damit beginnt das eigentliche Schneiden, die Formierung des Schnittpalts und das Austreiben der Schmelze. Die folgenden Prozesse spielen dabei eine zentrale Rolle: Energiebilanz, Schmelzfilm, Gasströmung und Materialaustrieb. Als wichtigste Einflussfaktoren auf diese Prozesse nennen Hügel und Graf [HG09] die Eigenschaften des Laserstrahls sowie die Fokussierung, die Eigenschaften von Gas und Schneiddüse und die Eigenschaften des Werkstoffs. Da es in dieser Arbeit um eine datenbasierte Annäherung an den Laserschneidprozess geht und nicht um den Versuch, den bestehenden

Modellen ein weiteres physikalisch motiviertes hinzuzufügen, soll hier – um ein Grundverständnis zu schaffen – nur kurz auf die Energiebilanz eingegangen werden. Für detailliertere Informationen zu allen Teilprozessen und Einflussgrößen sei auf die Literatur verwiesen [HG09; SM10; Yil17].

Globale Energiebilanz Die Energieeinkopplung, also die Umwandlung von Laserenergie in Prozesswärme, spielt beim Laserschneiden eine zentrale Rolle. Anhand der globalen Energiebilanz

$$P_a + P_{ch} - P_V = v s d \rho (c \delta_T + h_S) \quad (2.1)$$

lässt sich abschätzen, welche Blechdicke mit welcher Geschwindigkeit getrennt werden kann [HG09]. Auf der linken Seite der Gleichung steht die verfügbare Prozessleistung. Diese setzt sich zusammen aus der an der Schnittfront absorbierten Leistung P_a , welche der Leistung des Laserstrahls abzüglich der an der Oberfläche des Werkstücks und an der Schnittfront reflektierten und der durch den Schnittspalt transmittierten Leistung entspricht. Welcher Anteil der einfallenden Leistung reflektiert oder transmittiert wird, hängt insbesondere von der Neigung der Schnittfront ab. Diese nimmt mit wachsender Geschwindigkeit zu [Moh94]. Beim Brennschneiden mit Sauerstoff als Schneidgas wird die zur Verfügung stehende Leistung durch die chemische Verbrennungsleistung P_{ch} vergrößert. Reduziert wird sie durch Verluste P_V , aufgrund in das Werkstück fließender Wärme, thermischer Strahlung und konvektiver Wärmeabfuhr.

Auf der rechten Seite der Gleichung wird das pro Zeit aufschmelzbare Volumen (als Produkt der Schneidgeschwindigkeit v , der Schnittspaltbreite s und der Blechdicke d) multipliziert mit dem spezifischen Energiebedarf. Dieser setzt sich zusammen aus Dichte des Werkstücks ρ , der spezifischen Wärmekapazität des Materials c , der für das Schmelzen erforderlichen Temperaturdifferenz δ_T und der spezifischen Schmelzenthalpie h_S . Diese gibt an, wie viel Energie pro Masse für den Übergang vom festen zum flüssigen Aggregatzustand nötig ist. Je kleiner der spezifische Energiebedarf des Werkstoffs ist, desto größer ist das pro Zeit aufschmelzbare Volumen. Neben dem Aufschmelzen des Materials durch Einkopplung einer ausreichend großen Leistung sollte für einen kontinuierlichen Trennvorgang ein quasi-stationärer Zustand zwischen geschmolzenem und ausgetriebenem Material vorliegen.

2.1.3 Laserschneidmaschine

Abbildung 2.2 zeigt eine moderne 2D-Laserschneidmaschine. Als Strahlquelle dient bei dieser Maschine ein Festkörper-Scheibenlaser. Über das Lichtleitkabel wird die Strahlung vom Laser zum Schneidkopf innerhalb der Maschine transportiert. Dieser ist in Abbildung 2.3 zu sehen. Am oberen Ende des Schneidkopfs wird die Strahlung aus dem Lichtleitkabel ausgekoppelt. Dann wird sie mithilfe mehrerer beweglicher Optiken fokussiert, die Fokusposition und der Fokusdurchmesser werden eingestellt. Am unteren Ende transmittiert der Laserstrahl durch das Schutzglas, das die Optiken vor Verschmutzungen schützt, und verlässt den Schneidkopf durch die Düse. Die Maschine ist mit einer sogenannten fliegenden Optik ausgestattet. Das bedeutet, dass sich der Schneidkopf bewegt und das Blech in Ruhe auf den Stegen der Palette liegt.

Um die Maschine zu beladen, wird das Blech außerhalb auf die Palette gelegt. Danach wird es mit dem Palettenwechsler ins Innere der Maschine gefahren und die Bearbeitung kann beginnen. Kleine geschnittene Teile fallen zwischen den Stegen hindurch auf ein Förderband, mit dem sie aus der Maschine transportiert werden. Große Teile bleiben auf der Palette liegen und können nach dem Schneiden außerhalb der Maschine automatisiert oder manuell aus dem Restgitter entnommen werden. Die Stege der Maschine, auf deren Spitzen das Blech aufliegt verlaufen in y -Richtung und haben einen Abstand von 67 mm in x -Richtung und eine Breite von 3 mm. Die Stege sind keine durchgängigen Bleche, sondern haben – ähnlich wie ein Rechen – einzelne Spitzen, auf denen das Blech aufliegt.

Die Maschine ist wegen der gefährlichen Laserstrahlung (Laserklasse 4 [Deu15]) und den gesundheitsschädlichen Stäuben und Dämpfen komplett eingehaust und mit einer Absaugung, die zu einem Entstauber führt, versehen.

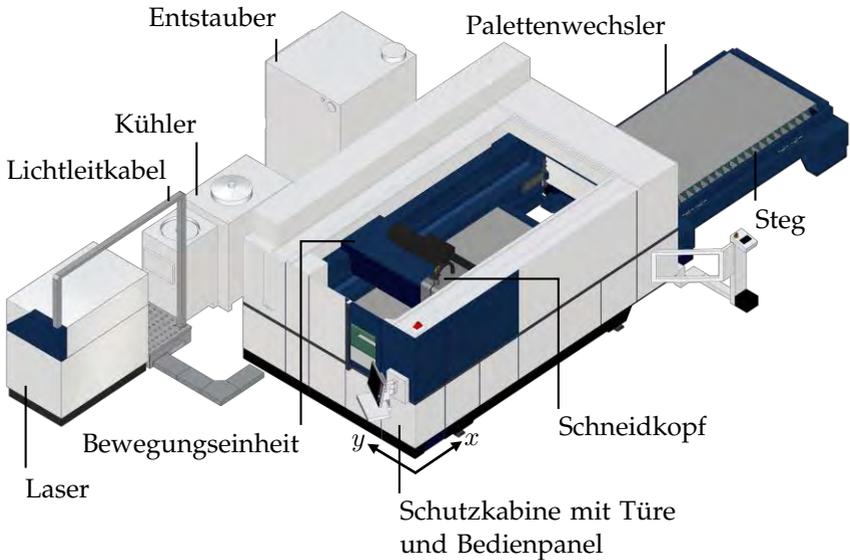


Abbildung 2.2 Moderne 2D-Flachbett-Laserschneidmaschine mit Festkörperlaser und fliegender Optik. Bei einem Arbeitsbereich von 3 m in x -Richtung und 1,5 m in y -Richtung hat die Maschine eine Aufstellfläche von 8,8 m \times 4,8 m. Quelle: TRUMPF.

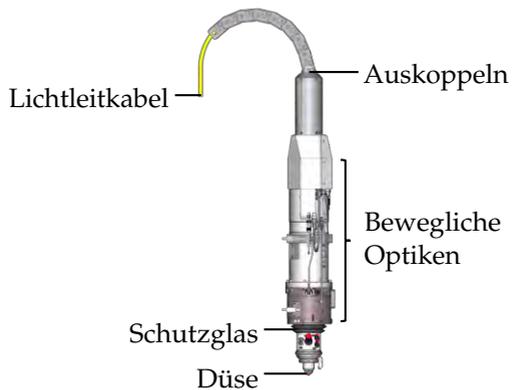


Abbildung 2.3 Schneidkopf der in Abbildung 2.2 gezeigten Laserschneidmaschine. Quelle: TRUMPF.

2.2 Qualität thermischer Schnitte

Das Prozessergebnis wird neben der erreichten Produktivität durch die Qualität des geschnittenen Teils definiert.

2.2.1 Schnittkantenqualität nach Norm

Hügel und Graf [HG09] nennen die folgenden Qualitätsaspekte beim Laserschneiden: Maßhaltigkeit, Schnittspaltbreite, Form der Schnittflanke, Rauigkeit, Bartbildung, Wärmeeinflusszone und Oxidfreiheit. Sie stellen fest, dass die Bearbeitungsqualität „keine absolut und für sich allein zu bewertende Größe“ ist, sondern von verschiedenen Randbedingungen, z. B. nachfolgenden Fertigungsprozessen, abhängt.

Die Norm DIN EN ISO 9013 [Deu17] definiert die Qualität eines thermischen Schnittes auf Basis der Form- bzw. Lagetoleranzen und des Aussehens der Schnittkante. Sie ist gültig für Laserstrahlschnitte in Materialien mit Dicken zwischen 0,5 und 32 mm. Da in dieser Arbeit nicht ganze geschnittene Teile, sondern nur gerade Schnittkanten betrachtet werden, können die Form- bzw. Lagetoleranzen nicht zur Bewertung hinzugezogen werden. Stattdessen wird die Prozessqualität anhand der Schnittkante beurteilt.

Die Qualität der Schnittfläche wird durch die beiden obligatorischen Größen Rechtwinkligkeits- oder Neigungstoleranz u und gemittelte Rautiefe $Rz5$ beschrieben. Bei ersterer handelt es sich um den Abstand zweier paralleler Geraden, die (bei Senkrechtschnitten) orthogonal zur Oberseite des Werkstücks verlaufen und zwischen denen das Schnittflächenprofil liegt. Dies ist in Abbildung 2.4(a) veranschaulicht. Die Messung von u wird in einem eingeschränkten Bereich der Schnittfläche durchgeführt, damit das Ergebnis nicht durch die Anschmelzung der Oberkante oder Gratbildung an der Unterkante beeinflusst wird. Im Falle eines 3 mm dicken Werkstücks muss der zu vermessende Bereich beispielsweise oben und unten jeweils um 0,3 mm verringert werden. Die gemittelte Rautiefe $Rz5$ ist das arithmetische Mittel der Differenzen zwischen der höchsten Spitze und der tiefsten Senke in fünf benachbarten Einzelmessstrecken eines Rauheitsprofils (siehe Abbildung 2.4(b)). Details zum Messgerät und zur Messung der Rautiefe sind in DIN EN ISO 3274 [Deu98a] beziehungsweise DIN EN ISO 4288 [Deu98b] aufgeführt und im folgenden Abschnitt

Tabelle 2.1 Qualitätsklassen für die Rechtwinkligkeitstoleranz u und Rautiefe $Rz5$ nach DIN EN ISO 9013 [Deu17] in Abhängigkeit von der Blechdicke d (in mm).

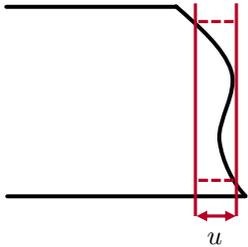
Klasse	u maximal in μm	$Rz5$ maximal in μm
1	$50 + (3 \cdot d/\text{mm})$	$10 + (0,6 \cdot d/\text{mm})$
2	$150 + (7 \cdot d/\text{mm})$	$40 + (0,8 \cdot d/\text{mm})$
3	$400 + (10 \cdot d/\text{mm})$	$70 + (1,2 \cdot d/\text{mm})$
4	$800 + (20 \cdot d/\text{mm})$	$110 + (1,8 \cdot d/\text{mm})$
5	$1200 + (35 \cdot d/\text{mm})$	–

2.2.1.1 kurz zusammengefasst. Rechtwinkligkeitstoleranz und Rautiefe müssen dort gemessen werden, wo die größten Werte zu erwarten sind, wobei es sich um zwei verschiedene Stellen handeln kann. Einzelne Fehler, wie z. B. sogenannte Kolkungen, müssen nicht einbezogen werden. Die Qualitätsklassen für Rechtwinkligkeitstoleranz und Rautiefe sind in Abhängigkeit von der Dicke des Werkstoffs d in Tabelle 2.1 gezeigt. Die Praxis hat gezeigt, dass beim Laserschneiden von rostfreiem Stahl einer Dicke von 3 mm mit einem Festkörperlaser die u -Qualitätsklasse 1 und die $Rz5$ -Qualitätsklasse 2 erreicht werden kann. Zusätzlich zu u und $Rz5$ können der Rillennachlauf, der projizierte Abstand zweier Schnittrillen in Schneidrichtung (siehe Abbildung 2.4(c)), und der Radius der Schnittoberkante (siehe Abbildung 2.4(d)) sowie der Grat an der Unterseite des Werkstücks zur Bewertung der Schnittkante hinzugezogen werden. Wie genau der Grat quantifiziert werden soll, ist nicht festgelegt.

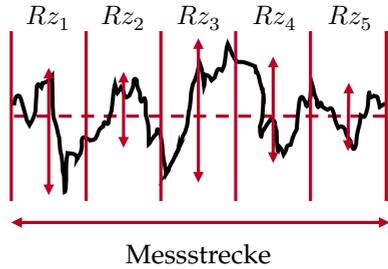
2.2.1.1 Details zur Rautiefenmessung

Die Ermittlung der Rautiefe basiert auf dem Rauheitsprofil der Schnittfläche. Dazu muss diese mit einem Tastschnittverfahren zweidimensional erfasst werden. Das tatsächlich gemessene Oberflächenprofil wird als ungefiltertes Primärprofil bezeichnet. Durch dessen Filterung mit einem Lang- bzw. Kurzpass-Gauß-Filter entstehen das relativ langwellige Welligkeitsprofil und das relativ kurzwellige Rauheitsprofil. Die drei unterschiedlichen Profile sind in Abbildung 2.5 gezeigt.

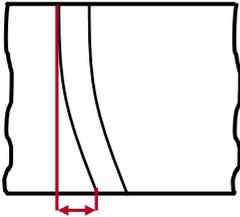
Der Übergang zwischen Welligkeits- und Rauheitsprofil wird durch die sogenannte Grenzwellenlänge definiert. Deren Wahl wiederum hängt von



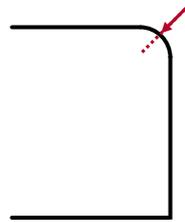
(a) Rechtwinkligkeitstoleranz u .



(b) Gemittelte Rautiefe $Rz5 = \frac{1}{5} \sum_i Rz_i$.



(c) Rillennachlauf.



(d) Anschmelzung der Oberkante.

Abbildung 2.4 Qualitätskenngrößen für eine thermische Schnittfläche nach DIN EN ISO 9013 [Deu17]. Die Schnittkante ist hier aus verschiedenen Perspektiven zu sehen. In den Abbildungen (b) und (c) blickt man frontal auf die Schnittkante (so wie bei den in dieser Arbeit gezeigten RGB-Aufnahmen). Bei den Abbildungen (a) und (d) ist die Blickrichtung dagegen parallel zu Schneidrichtung und Schnittfläche.

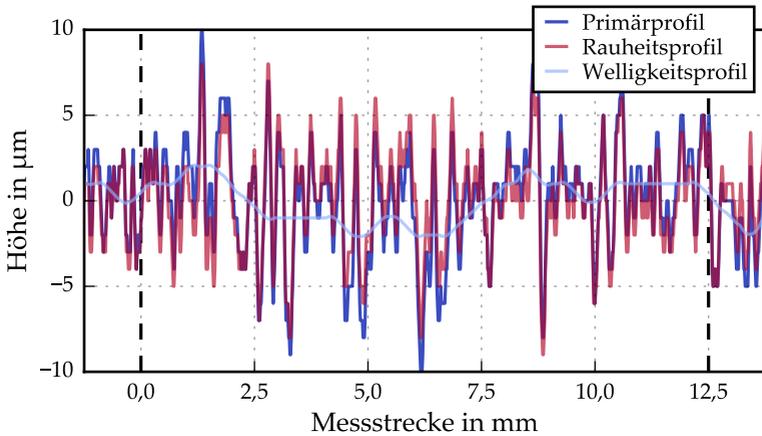


Abbildung 2.5 Mittels Filterung werden das Rauheits- bzw. das Welligkeitsprofil aus dem gemessenen Primärprofil extrahiert. Die Grenzwellenlänge für das Kurz- bzw. das Langpassfilter beträgt hier 2,5 mm; die ausgewertete Messstrecke ist 12,5 mm lang.

der erwarteten Rautiefe ab. Dies ist in Tabelle 2.2 erklärt. Beispielsweise gilt bei einer erwarteten Rautiefe von $40\ \mu\text{m}$ eine Grenzwellenlänge von 2,5 mm. Mit der Wahl der Grenzwellenlänge ist automatisch auch die Länge der Gesamtmessstrecke festgelegt. Ergibt sich nach Filterung des Primärprofils und Ermittlung der Rautiefe aus dem Rauheitsprofil ein Wert kleiner als 10 oder größer als $50\ \mu\text{m}$, müssen die Messdaten nochmals mit einer kleineren bzw. größeren Grenzwellenlänge ausgewertet werden.

2.2.2 Nachbearbeitung

Laserschneiden findet stets innerhalb einer Prozesskette statt. Vorgelagerte Prozesse sind beispielsweise die Erstellung eines Angebots oder die Generierung eines ausführbaren Programms für die Maschinensteuerung. Nachgelagerte Prozesse sind – je nach Produkt – Biegen, Zerspanen, Schweißen, Markieren, Stanzen, Beschichten, Verzinken, Lackieren und die Montage. Jeder dieser nachgelagerten Prozesse stellt bestimmte Anforderungen an die Qualität des geschnittenen Teils. Im Idealfall sollte eine Schnittkante keinen Grat sowie keine Spritzer- oder Schmelzperlen an der Ober- und Unterseite aufweisen. Sie sollte oxidfrei, möglichst rechtwinklig

Tabelle 2.2 Messung der Rautiefe nach DIN EN ISO 3274 [Deu98a] und DIN EN ISO 4288 [Deu98b]. Je nach erwarteter Rautiefe müssen die Grenzwellenlänge des Kurzpassfilters (zur Unterdrückung der Welligkeit) und die zugehörige Messstrecke angepasst werden.

$Rz5$ in μm	Grenzwellenlänge in mm	Messstrecke in mm
unter 0,1	0,08	0,4
0,1 bis 0,5	0,25	1,25
0,5 bis 10	0,80	4,0
10 bis 50	2,50	12,5
50 bis 200	8,00	40

und glatt sein [Sti19]. Oft sind scharfe Kanten nicht gewünscht, da sie die Verletzungsgefahr erhöhen und Probleme beim Beschichten verursachen. In vielen Produktionsanlagen stehen Maschinen oder Handwerkzeuge zur nachträglichen Verbesserung der Schnittkante bereit. Dies geschieht durch Entgraten oder Schleifen. Moderne Entgratmaschinen korrigieren störende Eigenschaften in mehreren Schritten mechanisch. Dazu verwenden sie verschiedene Schleifbänder, -bürsten und -walzen. Da jeder Nachbearbeitungsschritt die nötige Zeit und die Kosten pro Teil erhöht, sollte die Schnittqualität idealerweise so gut sein, dass keine Nachbearbeitung der Kanten und Teile nötig ist.

2.3 Modellierung des Laserschneidprozesses

Im Folgenden wird der Stand der Wissenschaft der Modellierung des Laserschneidprozesses erläutert. Dabei wird unterschieden zwischen physikalisch motivierten Ansätzen und datenbasierten Methoden des maschinellen Lernens. Bei der datenbasierten Modellierung versucht man, Zusammenhänge zwischen den Ein- und Ausgangsgrößen eines Systems ohne die Kenntnis des physikalischen Verhaltens abzubilden. Ein Modell wird dazu auf einem repräsentativen Datensatz, der möglichst alle Aspekte des abzubildenden Verhaltens enthält, trainiert. Anschließend wird das Modell zu Validierungszwecken auf Daten angewendet, die nicht zum Training verwendet wurden. Zu den datenbasierten Methoden zählen z. B.

neuronale Netze, evolutionäre Algorithmen und Fuzzy-Regel-basierte Systeme [SSA08]. Die Grundlagen des maschinellen Lernens und die in dieser Arbeit verwendeten Modellansätze sind in den Kapiteln 3 bzw. 5 nachzulesen.

2.3.1 Physikalisch motivierte Modellierung

Beim Laserschneiden sind mehrere physikalische Teilprozesse involviert. In analytischen Modellen wird meist nur ein Teilprozess betrachtet und das thermische Verhalten modelliert. Ein Modell wurde mit Gleichung (2.1) bereits vorgestellt. Publiziert wurden unter anderem stationäre thermische Modelle [CM87; Yil17] und solche, die die Bewegung des Laserstrahls einbeziehen [CA77; Ros46]. Außerdem wurden die analytischen Zusammenhänge zwischen verschiedenen Prozessparametern (z. B. Laserleistung) und der maximal trennbaren Blechdicke hergeleitet [KRL97]. Ivarson et al. [Iva+94] haben die Entstehung der Riefen auf der Schnittkante beim Brennschneiden in Abhängigkeit des Sauerstoffpartialdrucks, der Viskosität und der Oberflächenspannung der Schmelze betrachtet. Yilbas und Aleem [YA06] untersuchten den Schmelzfilm und die Entstehung von Gratperlen bei verschiedenen Laserleistungen und Gasdrücken beim Brennschneiden. Auch wenn es die analytischen Modelle ermöglichen, ein tieferes Verständnis eines einzelnen Teilprozesses unter ganz bestimmten Bedingungen zu erlangen, sind diese Modelle nicht in der Lage, den realen Gesamtprozess abzubilden.

Steen und Mazumder [SM10], Yilbas [Yil17] und Otto et al. [Ott+11] geben einen Überblick über numerische Modellierungsansätze. Da viele davon Jahrzehnte alt sind, wird hier nur exemplarisch auf einige aktuellere Arbeiten eingegangen.

Otto et al. [Ott+11] präsentieren einen numerischen Ansatz zur dynamischen Modellierung verschiedener Lasermaterialbearbeitungsprozesse. Abgebildet werden die Propagation des Laserstrahls zum Werkstück, die Interaktion des Strahls mit dem Material, Wärmeleitung und Konvektion sowie der Schmelzfluss. Durch die Simulation können die Schmelz- und Dampfströmung in Abhängigkeit von der Zeit visualisiert werden. Es werden nicht alle physikalischen Wechselwirkungen berücksichtigt, beispielsweise fehlt die Kopplung von Strömungs- und Thermodynamik.

Das Fraunhofer-Institut für Lasertechnik (ILT) in Aachen hat die Simulationssoftware QuCUT entwickelt. Bei dem zugrunde liegenden Modell handelt es sich um einen hybriden numerischen Ansatz. Mit QuCUT lässt sich der Zusammenhang zwischen fünf Designparametern einer Laseroptik (Strahlqualität, Astigmatismus, Fokusposition und Strahlradius in x- und y-Richtung) und der Riefenstruktur einer Schnittkante modellieren. Weitere Prozessparameter können nicht berücksichtigt werden. [PSS10]

Die Simulationssoftware CALCut [Pet16] vereint die Vorteile der analytischen und numerischen Ansätze zu einem dreidimensionalen stationären Modell. Es handelt sich dabei um eine Kombination verschiedener Algorithmen, mit der sich alle relevanten physikalischen Teilprozesse abbilden lassen. Als Eingangsgrößen dienen die Materialart und -dicke, Eigenschaften des Laserstrahls, Gasart und -druck sowie die Schneidgeschwindigkeit. Unter Berücksichtigung der räumlichen Verteilung der absorbierten Leistung, der Temperatur, der Dicke und Fließgeschwindigkeit des Schmelzfilms, der Verdampfungsrate und der Druckgradienten werden die Geometrien von Schnittfront und Schnittpalt, sowie die maximal mögliche Schneidgeschwindigkeit berechnet. CALCut ist sehr nützlich für die Erweiterung des Prozessverständnisses. Für die Unterstützung in der Praxis ist die Software jedoch nur bedingt geeignet, da über die Geometrien der Schnittfront und des Schnittpalts zwar Rückschlüsse auf die Qualität möglich sind, aber der Zusammenhang zwischen Eingangsgrößen und Qualitätskriterien der Schnittkante nicht direkt abgebildet wird.

2011 schrieben Otto et al. [Ott+11], dass Simulationen des Laserschneidprozesses in der industriellen Produktion nicht Stand der Technik sind. Daran hat sich bis heute wenig geändert. Trotz vieler Bemühungen in der Vergangenheit ist es bislang nicht gelungen, ein Modell zu entwickeln, das schnelle Antworten auf praxisrelevante Fragestellungen liefert. Als Hauptgrund dafür identifizieren Otto et al. [Ott+11] die enorme Komplexität des Prozesses durch die Wechselwirkung verschiedenster physikalischer Effekte (mehr dazu in Abschnitt 2.1). In den meisten Fällen werden deshalb stark vereinfachende Annahmen getroffen, die wiederum dazu führen, dass nur einzelne Phänomene beschrieben werden können. Datenbasierte Methoden könnten diese Problematik lösen, aufgrund ihrer Fähigkeit komplexe Zusammenhänge abzubilden. Sie sind dabei nicht als Ersatz, sondern eher als Ergänzung zu bestehenden Modellen zu sehen.

2.3.2 Datenbasierte Modellierung

Obwohl komplexe ML-Modelle, wie z. B. tiefe neuronale Netze, bereits erfolgreich in einigen Bereichen der fertigen Industrie eingesetzt werden (siehe Abschnitt 3.2.5), finden sie beim Laserschneiden bislang nur wenig Anwendung.

Einfache ML-Modelle wurden beispielsweise von Tercan et al. [Ter+16] eingesetzt. Sie verwendeten Klassifikations- und Regressionsbäume, um den Zusammenhang zwischen den Prozessparametern Strahlqualität, Astigmatismus, Fokusposition und Strahlradius und dem Fehlschnittisiko beziehungsweise der Rauheit der Schnittkante zu modellieren. Für das Training der Modelle wurden mit QuCUT [PSS10] generierte Simulationsdaten verwendet. Mithilfe dieser Modelle lassen sich Parameterfenster guter Qualität identifizieren und Zusammenhänge interpretieren [Ter+16]. Nachfolgearbeiten [Ter+17] verwendeten dieselbe Datengrundlage. Schnittkanten wurden anhand ihrer Rauheit mittels k-means Clustering in fünf verschiedene Qualitätsgruppen aufgeteilt. Die Gruppe mit den kleinsten Rauheitswerten wurde als gut definiert, die anderen als schlecht. Anschließend wurden wieder Klassifikationsbäume eingesetzt, um die Abhängigkeiten zwischen den Prozessparametern und guter bzw. schlechter Qualität zu modellieren.

Madić und Radovanović [MR12b] betrachteten ebenfalls den Zusammenhang zwischen verschiedenen Prozessparametern und einigen Qualitätskriterien. Sie übergaben die Prozessparameter Schneidgeschwindigkeit, Laserleistung, Gasdruck und Fokusposition an ein neuronales Netz, welches die Größe der Wärmeeinflusszone vorhersagte. Ihre experimentell erzeugte Datenbasis bestand aus 27 Schnitten in einem 3 mm dicken Edelstahlblech. Das verwendete neuronale Netz hatte vier Eingangs-, vier verdeckte und ein Ausgangsneuron. (Mehr zu diesen Begriffen steht in Abschnitt 3.2.1). Madić und Radovanović [MR12c] modellierten außerdem den Einfluss derselben Prozessparameter (ohne die Fokusposition) auf die Rauheit der Schnittkante anhand eines neuronalen Netzes. Die Datenbasis beinhaltete 25 Schnitte in 2 mm dickem Baustahl. Das trainierte neuronale Netz hatte drei Eingangs-, fünf verdeckte und ein Ausgangsneuron.

Tsai et al. [TLC08] haben ähnliche Untersuchungen für das Laserschneiden von Chipgehäusen angestellt. Sie modellierten den Zusammenhang zwischen drei Prozessparametern eines mit Dioden gepumpten Lasers

(Strom, Frequenz, Schneidgeschwindigkeit) und sechs Qualitätskenngrößen, z. B. der Breite der Wärmeeinflusszone, mit einem neuronalen Netz. Dieses wies eine verdeckte Schicht aus drei Neuronen auf. Die Datenbasis enthielt 27 Trainings- und 14 Testbeispiele. Um anschließend die optimalen Parameter hinsichtlich der Wärmeeinflusszone und der Geschwindigkeit zu finden, verwendeten sie einen genetischen Algorithmus. Das neuronale Netz diente dabei zur Berechnung der Fitnessfunktion.

Anstatt eines genetischen Algorithmus zur Optimierung der Prozessparameter für das Laserschneiden einer Aluminiumlegierung verwendeten Chaki et al. [CBB20] eine Partikelschwarmoptimierung basierend auf einem neuronalen Netz mit sieben verdeckten Neuronen. Das Netz schätzte vier Qualitätskenngrößen (z. B. Schnittspaltbreite) in Abhängigkeit von drei Prozessparametern (Schneidgeschwindigkeit, Pulsenergie und -breite). Die Datenbasis bestand aus 27 Schnitten mit verschiedenen Kombinationen der Prozessparameter.

Zusammenfassend ist zu sagen, dass es sich in allen genannten Fällen um sehr einfache Modellansätze handelt, die auf sehr kleinen Datenmengen (oder auf Simulationsdaten) entwickelt wurden. Wegen der jeweils verschiedenen Datengrundlagen und den wenigen Testdaten sind die Ergebnisse, insbesondere Angaben zur Modellgüte, nicht auf die vorliegende Arbeit übertragbar.

2.4 Laserschneiden in der Praxis

Diese Doktorarbeit wurde im Rahmen einer Industriepromotion bei der Firma TRUMPF Werkzeugmaschinen GmbH + Co. KG (im Folgenden TRUMPF genannt) angefertigt. Das Unternehmens-Portfolio umfasst Werkzeugmaschinen zum Biegen, Stanzen und für Laserschneid- und Laserschweißanwendungen. TRUMPF ist Weltmarktführer für (elektronische) Werkzeugmaschinen und industrielle Laser [Mül20; Sim07]. Da es keine Fachliteratur zum Laserschneiden in der Praxis gibt, kann hier nur die Sichtweise bei und im Umfeld von TRUMPF wiedergegeben werden. Die Kunden von TRUMPF sind überwiegend kleine bis mittelständische Unternehmen, die als Auftragsfertiger (engl. *job shopper*) arbeiten [Lei08]. Da sie keine eigenen Produkte anbieten, sondern Blechteile für andere Firmen fertigen, ändern sich Aufträge und Anforderungen häufig.

2.4.1 Beurteilung der Qualität

Bei TRUMPF sind Messgeräte zur Bestimmung der Rautiefe, Rechtwinkligkeitstoleranz und Grathöhe vorhanden. Wie in Abschnitt 1.1.1 bereits beschrieben, ist dies in kleinen Unternehmen meist nicht gegeben. Die Beurteilung der Qualität erfolgt deshalb subjektiv durch Einzelpersonen. Dabei wird ggf. zusätzlich einbezogen, welche Nachbearbeitungsprozesse verfügbar sind und wie das geschnittene Teil weiterverwendet werden soll.

2.4.2 Überprüfung des Maschinenzustands

Ein stabiler, klar definierter Zustand der Laserschneidmaschine ist die Grundvoraussetzung für die Kontrollierbarkeit des Prozesses, da nur in diesem Fall davon ausgegangen werden kann, dass die eingestellten Soll-Prozessparameter ausreichend gut mit den Ist-Prozessparametern übereinstimmen und Laserschneiden bis heute ein gesteuerter und kein geregelter Prozess ist. Eine Überprüfung des Maschinenzustands wird entweder regelmäßig, z. B. zu Schichtbeginn, oder beim Auftreten von Schneidproblemen durchgeführt.

Da eine automatische Überprüfung des Gesamtsystems nicht möglich ist und weil Menschen nicht in der Lage sind, vom Fehlerbild der Schnittkante auf die Ursache zu schließen (siehe Abschnitt 6.2.1.2), müssen verschiedene Einzelkomponenten nacheinander überprüft werden. Dazu zählen beispielsweise die Düse (Sauberkeit, keine Kratzer oder Dellen), die Strahlposition (der Laserstrahl sollte genau im Düsenmittelpunkt austreten) und das Schutzglas (Sauberkeit). Unterstützt wird der Bediener oder die Bedienerin dabei von Anleitungen im Benutzerhandbuch der Maschine oder in einer App sowie von Sensorik.

Die Probleme lassen sich am Beispiel des Schutzglases verständlich machen. Dieses befindet sich zwischen den optischen Elementen im Schneidkopf und der Düse, durch die der Strahl zusammen mit dem koaxialen Gasstrom austritt. Wie der Name schon sagt, dient das Schutzglas dem Schutz der Optiken vor Spritzern und Dämpfen. Ein verschmutztes Schutzglas hat einen verringerten Transmissionsgrad. Dies führt zu einem dazu, dass weniger Laserleistung zum Schneiden zur Verfügung steht und zum anderen dazu, dass sich das Schutzglas und der gesamte Schneidkopf

stärker erwärmen als üblich. Dadurch kann es zu einer thermisch induzierten Fokusverschiebung kommen, die sich auf die Schnittqualität auswirkt [TP20b].

Um die Sauberkeit des Schutzglases zu überprüfen, kann man es entweder ausbauen und mit bloßem Auge nach Verschmutzungen suchen, oder man verwendet die Schutzglas-Sensorik. Diese misst den Streulichtanteil am Rand des Glases orthogonal zur Strahlachse. Je mehr Schmutz vorhanden ist, umso stärker ist die Streuung. Beim Einsetzen eines neuen Schutzglases wird die Sensorik kalibriert. Bei späteren Messungen wird das detektierte Streulicht mit diesem Anfangswert verglichen und eine Abweichung in Prozent angegeben. In beiden Fällen muss der Mensch entscheiden, ab wann eine Verschmutzung so gravierend ist, dass das Glas gereinigt oder getauscht werden muss. Es ist schwierig einzuschätzen, wie groß die Auswirkung einer Verschmutzung dieser Komponente auf den Gesamtprozess ist. Außerdem erfordert diese Vorgehensweise eine geschulte Fachkraft und viel Zeit. Da ein Teil des Schneidkopfs für den Ausbau des Schutzglases abgenommen werden muss, kann es durch die Überprüfung zu weiteren Veränderungen im Gesamtsystem kommen.

Insbesondere bei voll automatisierten Maschinen, die ohne menschliche Eingriffe be- und entladen werden können, ist ein weiteres Problem, dass der Maschinenzustand nicht dauerhaft überwacht werden kann. Plötzliche Veränderungen können aber jederzeit auftreten, beispielsweise durch Spritzer oder durch Kollisionen des Schneidkopfs mit verkippten, hochstehenden Blechteilen. Unbemerkte Probleme, die die Schnittqualität verschlechtern, verursachen Ausschuss oder erhöhen den Nachbearbeitungsaufwand.

2.4.3 Ermittlung und Optimierung der Prozessparameter

Zusammen mit jeder Laserschneidmaschine werden Parametertabellen, die sogenannten LTTs (Laser-Technologie-Tabellen), angeboten. Diese beinhalten die einzustellenden Prozessparameter für verschiedene Materialien, Blechdicken und Bearbeitungsarten. Neben den Prozessparametern für unterschiedliche Schneidprozesse müssen auch passende Einstellungen für das Einstechen, das Beschriften usw. gefunden werden.

Die LTTs werden manuell erstellt. Normalerweise orientiert man sich dabei an ähnlichen Prozessen und versucht, von diesem Startpunkt aus

die Parameter experimentell zu optimieren. Sollen z. B. Parameter für eine höhere Laserleistung gefunden werden, beginnt man mit den Parametern für die nächst niedrigere Leistung. Die Optimierung der Prozessparameter basiert auf Bandbreitenvariationen der dominantesten Prozessparameter. Es werden maximal zwei Prozessparameter gleichzeitig variiert und mit jeder Kombination wird ein Teil geschnitten. So kann man sehen, wie sich welche Änderung auf die Schnittqualität auswirkt. Die Qualität der Schnitte wird anschließend manuell beurteilt. Dabei bezieht man gemessene Werte ein und vergleicht die erzielte Qualität mit Schnittkanten ähnlicher LTTs als Referenz. Entscheidend ist jedoch das subjektive Urteil des Versuchsexperten. Durch mehrere Bandbreitenvariationen versucht man, die optimale Parameterkombination iterativ zu finden.

Bei der Auswahl der Parameter muss stets ein Kompromiss eingegangen werden, da verschiedene Qualitätsziele (z. B. geringe Rautiefe und minimale Grathöhe) unvereinbar sein können. Gleichzeitig muss eine gewisse Robustheit des Prozesses sichergestellt werden. Dies realisiert man durch Gewährleistung einer Fokusbandbreite, was bedeutet, dass es bei Veränderungen der Fokusposition um beispielsweise 1 mm in jede Richtung nicht zu einer Veränderung der Schnittqualität kommen darf. Ebenfalls aus Stabilitätsgründen wird eine sogenannte Vorschubreserve von der maximal möglichen Schneidgeschwindigkeit abgezogen und so das Risiko für einen Fehlschnitt (d. h. eine Unterbrechung des Schnittspalts) verringert. Neben der Qualität und Stabilität des Prozesses stellt die Maximierung der Produktivität ein weiteres Optimierungskriterium dar. Bei der Ermittlung der LTTs muss man entscheiden, wie die drei genannten Optimierungsziele zu gewichten sind und Kompromisse eingehen. Durch die Vorschubreserve wird beispielsweise eine Verminderung der Schnittqualität und der Produktivität in Kauf genommen, um die Prozessstabilität zu erhöhen und das Fehlschnitttrisiko zu vermindern. Mit den neuen Parametern werden über mehrere Stunden und auf mehreren Maschinen verschiedene Teile geschnitten, um die Übertragbarkeit der Ergebnisse abzusichern. Anschließend werden einzelne, definierte Abnahmeteile hinsichtlich Rautiefe und Rechtwinkligkeitstoleranz vermessen.

Bei der Bestimmung neuer Schneiddaten verlässt man sich auf Einzelpersonen. Deren subjektive Einschätzung der Qualität, Gewichtung der Optimierungskriterien und Wissen über den Prozess sind entscheidend

für die resultierende LTT-Parameterkombination. Die damit verbundenen Nachteile wurden bereits beschrieben.

Oft ist es so, dass es trotz der gewissenhaften Ermittlung der LTT-Parameter beim Anwender zu Problemen kommt. Dafür gibt es drei Ursachen: Maschinenzustand, Materialtoleranzen und individuelle Qualitätspräferenzen. Treten Qualitätsprobleme auf, muss zuallererst der Maschinenzustand überprüft und gegebenenfalls korrigiert werden. Bestehen die Probleme weiterhin, ist dies meist dem Einfluss der Materialzusammensetzung und -qualität geschuldet, weil die LTTs in einem hochwertigen Referenzmaterial ermittelt werden und viele Firmen deutlich günstigere Bleche wechselnder Herkunft verwenden. Ein weiterer Grund für Probleme mit den LTTs ist, dass je nach Teil, späterem Einsatzzweck oder vorhandener Nachbearbeitung eine andere Gewichtung der Kriterien Qualität, Stabilität und Produktivität vorliegen kann. In beiden Fällen ist eine Anpassung der LTT-Parameter unumgänglich. Dazu sind unerfahrene oder schlecht ausgebildete Arbeitskräfte kaum in der Lage, was lange Maschinenstillstandszeiten, aufwendige Nachbearbeitung aller Teile oder Serviceeinsätze nach sich zieht. Da sich bei Auftragsfertigern sowohl Kunden als auch Anforderungen und Materialien häufig ändern, treten diese Probleme wiederholt auf.

Die Bearbeitung aller thematisierten Probleme in der Praxis würde den Rahmen dieser Arbeit sprengen. Der Fokus liegt auf der Bereitstellung der Rautiefe in der Produktion als objektives Qualitätskriterium (siehe Abschnitt 6.2.2), auf einer Lösung zur automatischen und zuverlässigen Überprüfung der Maschine (siehe Abschnitt 6.3) und auf Modellen, die den Zusammenhang zwischen den Prozessparametern und einzelnen Qualitätsmerkmalen abbilden und die bei der Ermittlung neuer LTTs zur Unterstützung eingesetzt oder gleich als Ersatz für die LTTs verwendet werden könnten (siehe Abschnitt 6.4.2, 6.4.3 bzw. 6.4.4). In Abschnitt 6.4.1 wird außerdem gezeigt, wie die Visualisierung von CNNs Unterstützung bei Schneidproblemen leisten könnte.

2.5 Diskussion des Stands der Technik

Der in diesem Kapitel beschriebene Stand der Technik sowie das Vorgehen in der Praxis bergen bereits ein großes Verbesserungspotential in sich. Wie

dringlich weitreichende Entwicklungen sind, wird bei der Betrachtung der veränderten und sich verändernden Anforderungen an die fertige Industrie deutlich.

Die fertige Industrie ist ein sehr etablierter Industriezweig. Ungefähr 9% aller Firmen und 28,5 Millionen Beschäftigte in der EU sind diesem Sektor zuzuordnen. Die generierte Wertschöpfung betrug 2017 1820 Milliarden Euro, was fast einem Drittel der gesamten in der Eurozone gewerblich erwirtschafteten Wertschöpfung entspricht. [Eur17]

Parallel zu den in Kapitel 1 genannten Herausforderungen befindet sich die fertige Industrie mitten in einem großen Veränderungsprozess, der im deutschen Sprachgebrauch mit Industrie 4.0 beschrieben wird. Im Englischen spricht man von *smart manufacturing* (dt. intelligente Fertigung). Diese Veränderung wird auch alle Unternehmen betreffen, die in der Lasermaterialbearbeitung tätig sind. Kusiak [Kus18] stellt zehn Vermutungen auf, wie die Fertigung der Zukunft aussehen könnte. Im Kontext dieser Arbeit sind die folgenden Punkte besonders relevant:

- Digitalisierung und damit zunehmende Abhängigkeit von Daten.
- Einsatz von Modellierungs- und Simulationswerkzeugen.
- Nutzung von Überwachungs- und Diagnosesysteme.

Sowohl im Fertigungsumfeld im Allgemeinen als auch bei anderen Lasermaterialbearbeitungsverfahren ist der Stand der Technik bezüglich der obigen drei Punkte weiter fortgeschritten als im Laserschneiden (siehe dazu auch Abschnitt 3.2.5). Für diese Diskrepanz könnte es verschiedene Gründe geben:

Datenverfügbarkeit und -generierung Komplexe ML-Modelle erfordern große Datenmengen. Alle in diesem Kapitel zitierten Arbeiten verwenden entweder Simulationsdaten [Ter+16; Ter+17] oder eine sehr geringe Anzahl selbst generierter experimenteller Daten [CBB20; MR12a; MR12b; MR12c; TLC08]. Während es beim Schweißen öffentlich verfügbare Daten gibt (z. B. GDXray [Mer+15]), ist die Existenz einer solchen Datenbank für das Laserschneiden nicht bekannt. Ein Grund für die kleinen Datensätze ist vermutlich, dass die Datengenerierung beim Laserschneiden sehr aufwendig ist, weil die Schnittkante während des Schneidens nicht zu

sehen ist und die Datenerfassung deshalb nachgelagert stattfinden muss. Die Zusammenführung der Informationen aus verschiedenen Datenquellen ist herausfordernd. Ähnliche Gründe könnten auch für das Fehlen ganzheitlicher Überwachungs- und Diagnosesysteme verantwortlich sein.

Interpretierbarkeit Mit den bislang verfügbaren Datensätzen und den eingesetzten Modellen lässt sich das große Potential von Methoden des maschinellen Lernens kaum nutzen. Komplexe Modellansätze, z. B. tiefe neuronale Netze, sind jedoch nur schwer zu interpretieren. Deshalb gibt es teilweise große Vorbehalte, diese Methoden einzusetzen. Die Tatsache, dass der Laserschneidprozess nicht vollständig verstanden ist, könnte in dieser Hinsicht besonders hinderlich sein, weil die Modellierung in einem solchen Fall auch zur Erweiterung des Verständnisses beitragen sollte. An der Interpretierbarkeit neuronaler Netze wird aber intensiv geforscht. Mehr dazu befindet sich in Abschnitt 3.2.6.

Unternehmensgröße Laserschneidmaschinen werden häufig von kleinen Firmen eingesetzt [Lei08], die als Auftragsfertiger oder Zulieferer arbeiten. Solche Unternehmen haben keine eigenen Entwicklungsabteilungen, die die Digitalisierung der Fertigung vorantreiben könnten. Da gerade kleinen Firmen ein Fachkräftemangel bevorstehen könnte [Mal+19], sind Modellierungs- und Diagnosesysteme besonders wichtig, weil sie die Bedienung vereinfachen und Unterstützung aus der Ferne ermöglichen.

Industrielle Reife Ein weiterer Punkt ist sicherlich, dass Laserschneiden seit vielen Jahrzehnten industriell eingesetzt wird und bereits eine große technische Reife erreicht hat.

3 Methoden des maschinellen Lernens

In diesem Kapitel werden zunächst die Grundlagen des maschinellen Lernens erklärt. Anschließend folgen eine ausführliche Einführung in künstliche neuronale Netze und ein Einblick in den Stand der Technik von Anwendungen dieser Modelle im Fertigungskontext.

3.1 Grundlagen

Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz. Der Begriff wurde 1959 von Samuel [Sam59] geprägt. Er schrieb, dass maschinelles Lernen Computern die Fähigkeit verleihe zu lernen, ohne explizit programmiert zu werden. Mitchell [Mit97] erklärte die Grundidee so: Ein Computerprogramm lerne aus einer Erfahrung E bezogen auf eine Aufgabe T und bewertet mit dem Leistungsmaß P , wenn die Leistung des Programms bei Aufgaben T (gemessen mit P) mit steigender Erfahrung E zunehme. Während bei der expliziten Programmierung aus der (Nicht-) Erfüllung einer Bedingung eine Aktion folgt, gilt beim maschinellen Lernen: Aus Daten entsteht ein Modell und aus diesem folgt eine Aktion. Mithilfe geeigneter Modellalgorithmen werden Muster oder Gesetzmäßigkeiten aus Lerndaten (von Mitchell [Mit97] als Erfahrung E bezeichnet) extrahiert. Anschließend kann das Modell auch für unbekannte Daten verwendet werden.

Obwohl das Konzept des maschinellen Lernens schon viele Jahrzehnte alt ist, hat dessen Popularität erst in den letzten Jahren stark zugenommen. Dies ist der zunehmenden Rechen- und Speicherkapazität und der Verfügbarkeit großer Datenmengen zu verdanken. Insbesondere in Bereichen, in denen eine explizite Programmierung schwierig ist, konnten damit große Erfolge erzielt werden (z. B. IBM-Watson [Hig12], DeepFace [Tai+14]).

Die Methoden des maschinellen Lernens lassen sich grob in drei Gruppen unterteilen: überwachtes, unüberwachtes und verstärkendes Lernen (engl. *reinforcement learning*). In dieser Arbeit kommt nur überwachtes Lernen zum Einsatz. Dabei wird eine Funktion aus Paaren von Ein- und Ausgangsdaten erlernt. Man weiß bereits (oder nimmt an), dass eine Beziehung zwischen Ein- und Ausgabe existiert. Der Algorithmus muss lediglich die Zusammenhänge entschlüsseln und abbilden. Beispiele für Algorithmen zum überwachten Lernen sind lineare Regression, Entscheidungsbäume und künstliche neuronale Netze (KNNs). [HTF09]

3.1.1 Wichtige Begriffe

Im Folgenden werden Grundbegriffe des überwachten maschinellen Lernens eingeführt.

3.1.1.1 Approximation einer Funktion

Beim überwachten maschinellen Lernen geht es darum, eine Funktion \hat{f} zu approximieren, die die allgemeine Form

$$\hat{y} = \hat{f}(\mathbf{x}) + \tau \tag{3.1}$$

hat. Dabei wird \mathbf{x} Input, Merkmal(e), Eingangs- oder unabhängige Variable(n) genannt, \hat{y} Output, Ausgangs- oder abhängige Variable. Der zufällige Fehlerterm τ hat einen Mittelwert von 0 und enthält z. B. nicht erfasste Variablen, die nützlich für die Vorhersage von y wären [Jam+13]. Die Funktion \hat{f} beinhaltet die systematische Information, die in \mathbf{x} über \hat{y} enthalten ist und ist meist unbekannt. Die Methoden des maschinellen Lernens fassen eine Vielzahl von Ansätzen zusammen, die alle das Ziel haben, den tatsächlichen Zusammenhang \hat{f} mithilfe einer Funktion f (dem Modell) zu approximieren. Das Ziel ist es also, eine Funktion f zu finden, für die gilt: $\hat{y} \approx f(\mathbf{x}) \forall \mathbf{x}$. Die zwei Hauptgründe für die Approximation von \hat{f} durch f sind die Vorhersage und die Erklärbarkeit.

Vorhersage In vielen Fällen sind die unabhängigen Variablen verfügbar, die abhängigen jedoch nicht. Dann kann f genutzt werden, um eine Vorhersage

$$y = f(\mathbf{x}) \quad (3.2)$$

für \hat{y} zu treffen. Die Genauigkeit der Vorhersage hängt von zwei Größen ab, dem reduzierbaren und dem nicht reduzierbaren Fehler [Jam+13]: Der reduzierbare Fehler hängt damit zusammen, dass f keine perfekte Approximation für \hat{f} ist. Dieser Fehler kann also reduziert werden, indem die Approximation verbessert wird, was beispielsweise durch die Verwendung flexiblerer Modelle erreicht werden kann. Diese bergen jedoch die Gefahr der Überanpassung (engl. *overfitting*) in sich (siehe Abschnitt 3.1.2.1). Der nicht reduzierbare Fehler wird durch die Abhängigkeit von \hat{y} vom Fehlerterm τ hervorgerufen, welcher nicht von \mathbf{x} abhängt. Er stellt die Grenze für die erreichbare Vorhersagegenauigkeit dar und ist in der Praxis normalerweise nicht bekannt.

Erklärbarkeit Oft möchte man verstehen, wie \hat{y} von \mathbf{x} abhängt. Zum Beispiel kann es interessant sein, welche unabhängigen Variablen für die Vorhersage relevant sind oder wie die Beziehung zwischen einer unabhängigen Variable und dem Output ist.

Kompromiss zwischen Genauigkeit und Interpretierbarkeit Je nachdem ob das Ziel der Approximation von \hat{f} das Treffen möglichst genauer Vorhersagen oder die Erklärbarkeit der Zusammenhänge oder beides ist, können unterschiedliche Approximationsmethoden vorteilhaft sein. Für die Verbesserung der Genauigkeit ist meist die Erhöhung der Modellkomplexität nötig, was die Interpretierbarkeit erschwert. In der Regel muss deshalb ein Kompromiss eingegangen werden zwischen der Genauigkeit der Vorhersage und der Interpretierbarkeit zur Erweiterung des Verständnisses [Jam+13].

3.1.1.2 Klassifikation und Regression

Je nachdem ob es sich bei der abhängigen Variable um eine quantitative oder qualitative (auch kategorische) handelt, liegt ein Regressions- oder ein

Klassifikationsproblem vor. Quantitative Variablen nehmen numerische Werte an, qualitative sind dagegen diskreten Klassen zuzuordnen. Oft lassen sich Probleme je nach Definition sowohl als Regressions- als auch als Klassifikationsaufgabe betrachten und viele Methoden des maschinellen Lernens können auf beide Fälle angewendet werden.

3.1.1.3 Training, Validierung und Test

Training Für die Approximation von \hat{f} mit f können unterschiedliche Methoden zum Einsatz kommen. Allen gemeinsam ist, dass die Anpassung mithilfe von N Beispieldatenpunkten geschieht und als Training des Modells bezeichnet wird. Die Beispieldaten werden deshalb als Trainingsdaten bezeichnet. Der Trainingsdatensatz $\{(\mathbf{x}_1, \hat{y}_1), (\mathbf{x}_2, \hat{y}_2), \dots, (\mathbf{x}_N, \hat{y}_N)\}$ besteht aus N Input-Output Paaren $(\mathbf{x}_n, \hat{y}_n)$ mit $n \in \{1, 2, \dots, N\}$, wobei $\mathbf{x}_n = (x_{n1}, x_{n2}, \dots, x_{nP})^\top \in \mathbb{R}^P$ der n -te Eingangsdatenpunkt und $\hat{y}_n \in \mathbb{R}$ der zugehörigen Output ist. Dieser wird auch Label (dt. Beschriftung) genannt.

Der nächste Schritt unterscheidet sich für parametrische und nicht parametrische Methoden. Bei letzteren werden keinerlei Annahmen bezüglich der Funktion f getroffen. Diese Methode kommt in der vorliegenden Arbeit nicht zum Einsatz und wird deshalb nicht weiter beschrieben. Bei den parametrischen Methoden wird im ersten Schritt eine Vermutung hinsichtlich der Funktion f aufgestellt, was das Problem sehr vereinfacht. Dafür muss man allerdings in Kauf nehmen, dass die angenommene Funktion f nicht perfekt zur tatsächlichen Form von \hat{f} passen könnte. Kann beispielsweise angenommen werden, dass es sich bei einer Regression um einen linearen Zusammenhang handelt, dann ist

$$f_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1x_1 + w_2x_2 + \dots + w_Px_P \quad (3.3)$$

die gesuchte Funktion [Jam+13]. Die Modellparameter werden auch als Gewichte bezeichnet. Nicht alle von ihnen spielen dieselbe Rolle. Geometrisch definieren die Gewichte (engl. *weights*) w_1, w_2, \dots, w_P die Winkel der aufgespannten Hyperebene bezogen auf die Koordinatenachsen. Der sogenannte Bias w_0 beschreibt dagegen eine Verschiebung der Hyperebene aus dem Koordinatenursprung [Kub15].

Analog dazu kommt bei Klassifikationsproblemen im binären Fall (also bei zwei Klassen $\{0, 1\}$) die logistische Regression zum Einsatz [Jam+13]:

$$f_{\mathbf{w}}(\mathbf{x}) = \begin{cases} 0, & \text{für } \text{pr} < 0,5 \\ 1, & \text{für } \text{pr} \geq 0,5 \end{cases} \quad (3.4)$$

$$\text{mit } \text{pr} = \frac{e^{(w_0 + w_1 x_1 + w_2 x_2 + \dots + w_P x_P)}}{1 + e^{(w_0 + w_1 x_1 + w_2 x_2 + \dots + w_P x_P)}}.$$

Hier gibt pr die Wahrscheinlichkeit dafür an, dass ein Datenpunkt \mathbf{x} zur Klasse $y = 1$ gehört.

Nach der Auswahl der Funktion müssen die Modellparameter \mathbf{w} mithilfe der Trainingsdaten dahingehend optimiert werden, dass gilt: $\hat{y} \approx f(\mathbf{x}) \forall \mathbf{x}$. Dafür gibt es verschiedene Methoden. Für jede Wahl der Parameter ergibt sich ein anderes Modell.

Kostenfunktion Für die Optimierung der Modellparameter wird eine Kosten- oder Verlustfunktion $E(\mathbf{w})$ (engl. *error function*) definiert, die minimiert werden soll. Diese Funktion beschreibt die Güte der Modellvorhersage, indem sie den Output des Modells mit dem jeweils gewünschten Output (d. h. dem Label) für jedes Trainingsbeispiel vergleicht. Es gibt verschiedene Möglichkeiten, die Kostenfunktion zu definieren. Ein für Regressionsaufgaben häufig gewählter Ansatz ist die Minimierung des über alle N Trainingsbeispiele gemittelten quadratischen Fehlers zwischen der Modellvorhersage y_n und dem Label \hat{y}_n . Die Kostenfunktion lautet dann [Jam+13]:

$$E(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^N \underbrace{(f_{\mathbf{w}}(\mathbf{x}_n) - \hat{y}_n)}_{=y_n}^2. \quad (3.5)$$

Weitere Metriken, die als Verlustfunktion eingesetzt werden können, sind in Abschnitt 3.1.2 zu finden.

Als Trainingsdaten für die Optimierung der Modellparameter werden nicht alle verfügbaren Daten verwendet, sondern nur eine Teilmenge. Diese wirken sich unmittelbar auf das Modell aus. Der vorhandene Datensatz wird in drei disjunkte Teilmengen aufgeteilt: Trainings-, Validierungs- und

Testdaten. Es gibt verschiedene Methoden, die Daten aufzuteilen. Zum Beispiel kann die Aufteilung zufällig erfolgen oder hinsichtlich bestimmter Verteilungen der vorhandenen Klassen. Üblicherweise beinhaltet der Trainingsdatensatz den größten Anteil der vorhandenen Daten, Validierungs- und Testdatensatz sind ungefähr gleich groß. Die Validierungs- und die Testdaten werden für die Evaluierung der Modellgüte mithilfe einer geeigneten Metrik verwendet [KJ13].

Validierung Die Validierungsdaten werden während des Entwicklungsprozesses eines Modells genutzt, um die Auswirkungen verschiedener Hyperparameter auf die Modellgüte zu bewerten. Bei den Hyperparametern handelt es sich um Parameter „höherer Ebene“, z. B. die Annahme eines quadratischen statt eines linearen Zusammenhangs oder die Anzahl der Schichten eines neuronalen Netzes. Die Validierungsdaten wirken sich deshalb (durch etwaige Anpassungen der Hyperparameter) indirekt auf das Modell aus.

Test Die Testdaten kommen erst dann zum Einsatz, wenn die Hyperparameter des Modells festgelegt und das Modell auf Basis der Trainingsdaten trainiert wurde [RND10]. Die Modellgüte wird damit auf unbekanntem Daten evaluiert, die keinerlei Einfluss auf das Modell hatten. Nur der Testfehler gibt eine verlässliche Auskunft über die Fähigkeit des Modells, Vorhersagen für unbekannte Eingangsdaten zu treffen.

Kreuzvalidierung Anstatt ein Modell auf einem Trainingsdatensatz zu trainieren und auf einem Validierungsdatensatz zu evaluieren, wird häufig die sogenannte Kreuzvalidierung verwendet, um eine verlässliche Aussage über die Modellgüte und die Sensitivität hinsichtlich der Aufteilung der Daten zu erhalten. Die bekannteste Methode ist die k -fache Kreuzvalidierung. Die zur Verfügung stehende Datenmenge (ohne Testdaten) wird dabei in k gleichgroße Teilmengen aufgeteilt. Anschließend werden k Durchläufe gestartet, bei denen jeweils $k - 1$ Teilmengen als Trainingsdaten verwendet werden und die verbleibende Teilmenge zur Validierung eingesetzt wird [Jam+13]. Anschließend werden der Durchschnitt und die Verteilung der Fehler der k Durchläufe berechnet.

3.1.2 Evaluierung der Modellgüte

Um die Güte eines Modells zu quantifizieren, um also zu beurteilen, wie gut die vom Modell getroffenen Vorhersagen zu den tatsächlichen Datenpunkten passen, wird eine geeignete Metrik benötigt. Die Metriken sind hier für den allgemeinen Fall eines Modells mit mehreren Ausgängen \mathbf{y} definiert. y_{nm} beschreibt dabei den Wert des m -ten Ausgangs für den n -ten Datenpunkt. Für Regressionsprobleme wird meistens der mittlere quadratische Fehler (engl. *mean squared error*)

$$\text{MSE} = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M (y_{nm} - \hat{y}_{nm})^2 \quad (3.6)$$

benutzt [Jam+13], mit der Anzahl der Datenpunkte (des Trainings-, Validierungs- oder Testdatensatzes) N und der Anzahl abhängiger Variablen M . Das Label oder der wahre Wert ist \hat{y}_{nm} , die Modellvorhersage ist y_{nm} .

Eine andere Möglichkeit, bei der einzelne große Fehler weniger stark ins Gewicht fallen, ist der mittlere absolute Fehler (engl. *mean absolute error*)

$$\text{MAE} = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M |y_{nm} - \hat{y}_{nm}|. \quad (3.7)$$

Im Falle von Klassifikatoren ist die Korrekt-Klassifikationsrate (engl. *classification accuracy*)

$$\text{CA} = 1 - \frac{1}{2} \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M |y_{nm} - \hat{y}_{nm}| \quad (3.8)$$

eine häufig verwendete Qualitätsmetrik. Sie beschreibt den Anteil der korrekten Klassenzuordnungen. Hier bezeichnet N ebenfalls die Anzahl der Datenpunkte, M die Anzahl der Klassen, \hat{y}_{nm} die wahre und y_{nm} die vorhergesagte Klasse. Bei mehreren Ausgängen erhält der wahrscheinlichste Ausgang bei der Vorhersage den Wert 1 und alle anderen den Wert 0 (engl. *one-hot encoding*) [Fri14]. Falls die wahre und die vorhergesagte Klasse übereinstimmen, verschwindet der Term innerhalb des Summenzeichens, ansonsten ist er 1. Warum durch 2 geteilt wird, lässt sich am

Beispiel mit zwei Ausgängen ($M = 2$) leicht verstehen: Wenn die Vorhersage eines Trainingsbeispiels nicht korrekt ist, ist sowohl $|y_{n1} - \hat{y}_{n1}| = 1$ als auch $|y_{n2} - \hat{y}_{n2}| = 1$. Damit dieser Fehler nicht doppelt eingeht, muss durch 2 geteilt werden. CA beschreibt den Anteil der korrekt zugeordneten Beispiele einer Klasse bezogen auf die Gesamtanzahl aller Beispiele im Datensatz. Bei unausgeglichene Anteilen der Klassen im Datensatz ist die gewichtete Korrekt-Klassifikationsrate (engl. *balanced classification accuracy*, BCA) aussagekräftiger. Die Anzahl der korrekt klassifizierten Beispiele einer Klasse wird dabei nur auf die Anzahl der Beispiele dieser Klasse im Datensatz bezogen. Anschließend wird über alle Klassen gemittelt. Bei einer binären Klassifikation, bei der 99 % der Datenpunkte zu Klasse 1 und 1 % zu Klasse 0 gehören, wäre die CA 99 %, wenn das Modell immer Klasse 1 vorhersagt. Die BCA wäre dagegen nur 50 %.

Zur Beurteilung der Klassifikationsgüte wird außerdem oft die sogenannte Wahrheitsmatrix verwendet [Jam+13]. Diese ist in Tabelle 3.1 für eine binäre Klassifikation zu sehen. Sie beinhaltet die Häufigkeiten für vier verschiedene Fälle: richtig positiv (rp), falsch negativ (fn), falsch positiv (fp) und richtig negativ (rn). Worum es sich dabei handelt, lässt sich am besten an einem Beispiel verstehen: Angenommen ein Modell soll anhand eines Bildes vorhersagen, ob ein Produkt einen Defekt hat (1) oder nicht (0). Richtig positiv sind alle Fälle, bei denen das Modell einen Defekt vorhersagt und die auch tatsächlich einen Defekt aufweisen. Zu den falsch negativen Vorhersagen werden alle Vorhersagen des Modells gezählt, bei denen kein Defekt erkannt wird, tatsächlich aber einer vorliegt. Bei den falsch positiven ist es gerade umgekehrt. Richtig negativ sind die Fälle, bei denen weder ein Defekt erkannt wird noch einer vorliegt. Die Korrekt-Klassifikationsrate lässt sich dann auch als Verhältnis von (rp + rn) zu (rp + fp + fn + rn) ausdrücken. Aus der Wahrheitsmatrix können außerdem die Precision (dt. Klassifikationsgenauigkeit), das Verhältnis von rp zu (rp + fp), und der Recall (dt. Sensitivität), das Verhältnis von rp zu (rp + fn), abgeleitet werden.

Eine weitere Möglichkeit, die Güte eines Klassifikators zu beschreiben, ist die kategorische Kreuzentropie (engl. *categorical cross-entropy*) [Mur12]:

$$\text{CCE} = -\frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M \hat{y}_{nm} \log(y_{nm}). \quad (3.9)$$

Tabelle 3.1 Die Wahrheitsmatrix für eine binäre Klassifikation beinhaltet die Häufigkeiten der vier Fälle: richtig positiv (rp), falsch negativ (fn), falsch positiv (fp) und richtig negativ (rn).

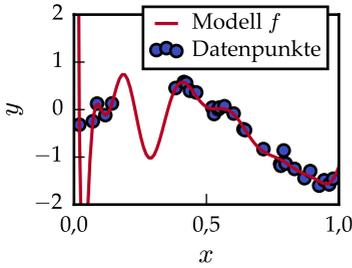
		Wahrheit	
		1	0
Vorhersage	1	rp	fp
	0	fn	rn

\log ist hier der natürliche Logarithmus. Während die Labels $\hat{y}_n m$ als *one-hot* Vektor $\hat{\mathbf{y}}_n$ vorliegen, beinhaltet der Vektor der Vorhersagen \mathbf{y}_n Werte $y_n m$ zwischen 0 und 1, die in Summe gerade 1 ergeben [Bis06]. Diese können als Wahrscheinlichkeiten der Klassen interpretiert werden. Die CCE nimmt stets positive Werte an, wobei 0 für ein perfektes Modell steht, welches immer (mit der Wahrscheinlichkeit 1) die korrekte Klasse vorhersagt.

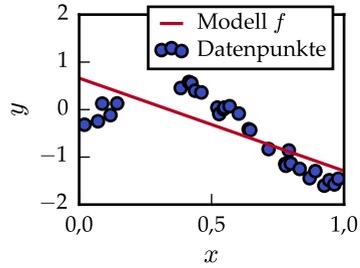
3.1.2.1 Bias-Varianz-Dilemma

Die Abweichung zwischen dem vom Modell vorhergesagten und dem wahren Wert hängt von drei fundamentalen Größen ab, dem nicht reduzierbaren Fehlerterm τ (siehe Formel (3.1)) sowie der Varianz und dem Bias (dt. Voreingenommenheit, Verzerrung) [Jam+13]. Das Bias-Varianz-Dilemma (engl. *bias-variance tradeoff*) beschreibt die Schwierigkeit der gleichzeitigen Minimierung dieser zwei Fehlerquellen.

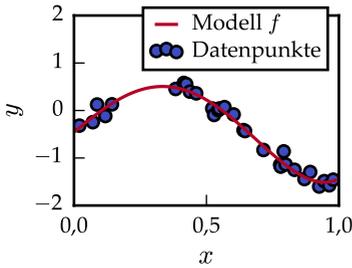
Die Varianz beschreibt den Einfluss einer Änderung der Trainingsdaten auf das Modell f . Ein Modell mit einer hohen Varianz reagiert empfindlich auf einzelne Datenpunkte. Anstatt die systematische Beziehung zwischen Ein- und Ausgang abzubilden, modelliert es das Rauschen in den Trainingsdaten. Dieses Problem wird auch als Überanpassung bezeichnet. Ob ein Modell mit hoher Varianz vorliegt, kann man durch einen Vergleich des Trainings- und des Validierungsfehlers herausfinden. Ist der Trainingsfehler deutlich kleiner als der Validierungsfehler, liegt Überanpassung vor. Ein Beispiel für ein überangepasstes Modell ist in Abbildung 3.1(a) zu sehen. Ein hoher Bias tritt auf, wenn ein komplexer realer Zusammenhang mit einem (vergleichsweise) einfachen Modell approximiert wird. Werden



(a) Zu hohe Varianz.



(b) Zu hoher Bias.



(c) Guter Kompromiss.

Abbildung 3.1 Beispiele für ein Modell mit zu hoher Varianz, zu hohem Bias und für ein Modell mit einem guten Kompromiss zwischen Bias und Varianz.

beim Aufstellen der Vermutung hinsichtlich des Zusammenhangs zwischen unabhängigen und abhängigen Variablen (siehe Abschnitt 3.1.1.3) falsche Annahmen getroffen und führen diese dazu, dass die Beziehungen unabhängig von der Anzahl der Trainingsbeispiele nicht richtig modelliert werden können, weist das Modell einen Bias auf. Dieses Problem wird auch als Unteranpassung bezeichnet. Ein Beispiel dafür ist in Abbildung 3.1(b) gezeigt. Ob Unteranpassung vorliegt, ist nicht leicht herauszufinden. Ein hoher Bias führt zu einem hohen Trainings- und einem ähnlich hohen Validierungsfehler. Die Beurteilung der Größe eines Fehlers hängt jedoch stark von der jeweiligen Situation und Aufgabenstellung ab.

Je nachdem ob ein Modell mit einem hohen Bias oder einer hohen Varianz vorliegt, sollten verschiedene Ansätze zur Problembewegung herangezogen werden. Um den Bias zu reduzieren, wird üblicherweise die Modellkomplexität erhöht [HTF09]. Da es dadurch zu einer Erhöhung

der Varianz kommen kann, muss stets ein Kompromiss gefunden werden. Die Reduzierung der Varianz kann man neben der Verringerung der Modellkomplexität beispielsweise auch durch das „rechtzeitige“ Beenden des Optimierungsprozesses oder durch die sogenannte Regularisierung im Trainingsprozess erreichen [BG11]. Dabei wird verhindert, dass der Einfluss einzelner Modellparameter zu groß wird, z. B. durch Hinzufügen eines weiteren Terms zur Kostenfunktion. Ein Beispiel für einen guten Kompromiss zwischen Bias und Varianz ist in Abbildung 3.1(c) zu finden.

3.1.3 Arbeitsablauf eines ML-Projektes

Der Arbeitsablauf eines ML-Projektes kann in fünf Phasen gegliedert werden, die gegebenenfalls mehrfach durchlaufen werden müssen [Pan19]:

- Daten generieren/sammeln
- Datenvorverarbeitung (siehe z. B. Abschnitt 4.2.2.1)
- Datenanalyse und Modellauswahl (inkl. Merkmalsdesign)
- Trainieren, Validieren und Testen des Modells
- Evaluierung.

Diese Phasen wurden auch bei der Erstellung der vorliegenden Arbeit durchlaufen. Wie sie miteinander und mit vor- und nachgelagerten Arbeitsschritten zusammenhängen, ist in Abbildung 3.2 veranschaulicht. Eine in der Praxis häufig angewandte Schätzung ist, dass ca. 80 % der Zeit für die Datenvorverarbeitung investiert werden müssen und nur ca. 20 % der Zeit für die Modellauswahl, das Training und die Evaluierung zur Verfügung stehen [Pan19]. Damit decken sich auch die Ergebnisse des CrowdFlower Data Science Reports [Cro16]. Dieser macht transparent, dass Datenwissenschaftler (engl. *data scientists*) durchschnittlich 78 % ihrer Arbeitszeit mit der Bereinigung, der Organisation und der Erfassung von Daten verbringen. 10 % der Zeit wenden sie für das Design geeigneter Trainingsdatensätze auf und nur 3 bzw. 4 % der Zeit verbleiben für die Analyse der Daten und die Entwicklung der Algorithmen. Es ist davon auszugehen, dass sich diese Anteile im industriellen Umfeld noch weiter verschieben, weil die Datengenerierung und -bereinigung dort oft sehr aufwendig sind.

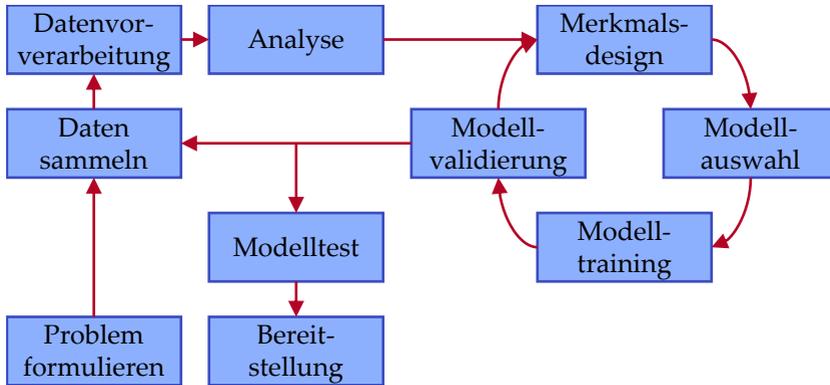


Abbildung 3.2 Arbeitsablauf eines ML-Projektes, beginnend mit der Formulierung der Problemstellung, frei nach Pant [Pan19] und Visengeriyeva et al. [Vis+21].

3.2 Künstliche neuronale Netze

Bei den parametrischen Methoden wird zunächst eine Vermutung f (genauer: f_w) hinsichtlich des systematischen Zusammenhangs zwischen Ein- und Ausgangsgrößen aufgestellt und dann werden die Parameter dieser Funktion bestimmt. Bislang wurden für f nur die beiden sehr einfachen Ansätze der linearen und der logistischen Regression in den Formeln (3.3) und (3.4) eingeführt. Eine andere Möglichkeit, die die Approximation komplexer Zusammenhänge ermöglicht, sind künstliche neuronale Netze.

Künstliche neuronale Netze wurden inspiriert von biologischen neuronalen Netzen und der Vernetzung von Neuronen im Nervensystem eines Lebewesens. Das Ziel bei ihrer Verwendung ist nicht, das biologische Vorbild exakt nachzubilden, sondern die Verarbeitung von Informationen und die Modellbildung. Künstliche neuronale Netze bestehen aus vielen kleinen Einheiten, den sogenannten Neuronen, die mithilfe gewichteter Verbindungen zu größeren Strukturen verbunden werden [Kub15]. Die Anfänge gehen auf McCulloch und Pitts [MP43] zurück, die bereits in den 1940er Jahren ein Netzwerk bestehend aus Verknüpfungen elementarer Einheiten vorstellten.

Neuronale Netze sind universelle Approximatoren (Universelles Approximationstheorem [Pin99]). Das bedeutet, dass damit jede stetige Funktion

beliebig genau approximiert werden kann und dass diese Modelle prinzipiell auf jedes beliebige Problem angewendet werden können. Das Universelle Approximationstheorem erklärt jedoch nicht, wie viele Neuronen dafür nötig sind und wie die Modellparameter gewählt werden müssen. Obwohl sicher ist, dass eine Lösung existiert, kann nicht garantiert werden, dass sie gefunden wird [Kub15].

3.2.1 Grundlagen künstlicher neuronaler Netze

Das Netz wird durch eine Funktion f beschrieben. Ein vorwärts gerichtetes Netz besteht aus L Schichten und jede Schicht besteht aus mehreren Neuronen. Die Werte der Neuronen werden als Aktivierungen bezeichnet. Bei $a_i^{(l)}$ handelt es sich um die Aktivierung des i -ten Neurons der l -ten Schicht. Der Input wird an die Neuronen der ersten Schicht $l = 1$ übergeben, $\mathbf{a}^{(1)} = \mathbf{x}$, und durch das Netz auf den Output in der letzten Schicht $l = L$ abgebildet: $\mathbf{y} = \mathbf{a}^{(L)}$. Die Schichten 2 bis $L - 1$ sind die sogenannten verdeckten Schichten (engl. *hidden layers*). [Kub15]

Für die Schichten gilt Folgendes: Die Aktivierungen der Schicht $\mathbf{a}^{(l+1)}$ werden mithilfe einer nicht linearen Transformation $H_{l,l+1}$ aus den Werten der davor liegenden Schicht $\mathbf{a}^{(l)}$ berechnet:

$$\mathbf{a}^{(l+1)} = H_{l,l+1} \left(\mathbf{a}^{(l)}, \mathbf{W}_{H_{l,l+1}} \right). \quad (3.10)$$

Bei $H_{l,l+1}$ handelt es sich oft um die Verkettung einer affinen Transformation von $\mathbf{a}^{(l)}$, die mit $\mathbf{W}_{H_{l,l+1}}$ parametrisiert ist, mit einer nicht linearen Aktivierungsfunktion [SGS15]. Die Eingangsinformation wird an die Neuronen der ersten Schicht übergeben und dann Schicht für Schicht über gewichtete Verbindungen weitergegeben bis zum Ausgang. Dieser Prozess wird als Vorwärts-Propagation (engl. *forward propagation*) bezeichnet.

3.2.1.1 Vollständig verbundene neuronale Netze

In einem vollständig verbundenen neuronalen Netz (engl. *fully connected*, FC) ist jedes Neuron a_j einer Schicht $l + 1$ mit jedem Neuron a_i der davor liegenden Schicht l über eine mit w_{ij} gewichtete Verbindung vernetzt. Ein

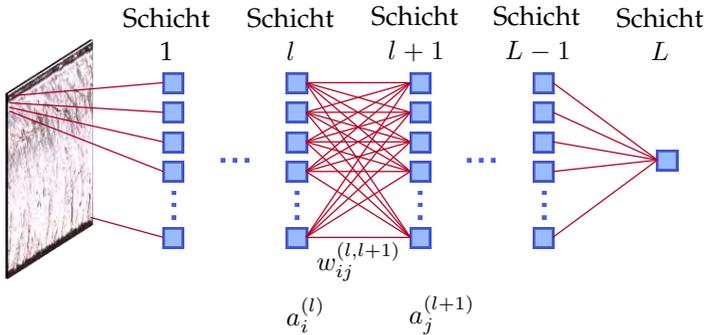


Abbildung 3.3 Vollständig verbundenes neuronales Netz. Als Eingangsinformation wird hier ein Bild übergeben. Dabei erhält jedes Neuron der ersten Schicht den Wert eines Pixels. Andere Eingangsdaten sind ebenfalls möglich. Die Information wird Schicht für Schicht durch das Netz propagiert (im Bild von links nach rechts), wobei die Verbindungen zwischen den Neuronen mit $w_{ij}^{(l,l+1)}$ gewichtet sind. Um die Aktivierung $a_j^{(l+1)}$ eines Neurons zu berechnen, kommt nach dem Aufsummieren aller gewichteten Aktivierungen der vorhergehenden Schicht l und dem (nicht eingezeichneten) Bias eine nicht lineare Aktivierungsfunktion zum Einsatz. Das gezeigte neuronale Netz hat einen Ausgang, mehrere sind genauso möglich.

solches Netz ist in Abbildung 3.3 zu sehen. Die allgemeine Form eines solchen Netzes ist [Mon+17]:

$$a_j^{(l+1)} = \text{act} \left(\sum_i a_i^{(l)} \cdot w_{ij}^{(l,l+1)} + b_j^{(l+1)} \right) \quad (3.11)$$

mit $a_j^{(1)} = x_j$ und $y_j = a_j^{(L)}$. Mit act wird die nicht lineare Aktivierungsfunktion beschrieben. Weitere Informationen dazu befinden sich im nächsten Abschnitt. Der sogenannte Bias-Term b_j ermöglicht eine Verschiebung der Aktivierungsfunktion. Er ist hier der Vollständigkeit halber aufgeführt, wird im Folgenden jedoch für eine einfachere Lesbarkeit häufig weggelassen. Sowohl b_j als auch die Gewichte w_{ij} werden während des Trainingsprozesses optimiert.

3.2.1.2 Aktivierungsfunktionen

Die Aktivierungsfunktion integriert eine nicht lineare Komponente in das neuronale Netz. Ohne diese könnte es den Eingang (unabhängig von der

Anzahl der Neuronen/Schichten) nur linear auf den Ausgang abbilden. Erst die Verwendung der Aktivierungsfunktion ermöglicht die Modellierung komplexer Zusammenhänge. Die Aktivierungsfunktion wird, wie in Formel (3.11) gezeigt, in jedem einzelnen Neuron auf die gewichtete Summe der Aktivierungen der Vorgängerschicht angewendet. Sie bestimmt, ob das Neuron aktiviert wird oder nicht. Es gibt verschiedene häufig verwendete Aktivierungsfunktionen mit unterschiedlichen Vor- und Nachteilen. Im Folgenden werden nur Funktionen eingeführt, die in dieser Arbeit verwendet werden.

Logistische Funktion Bei der logistischen Funktion handelt es sich um einen Spezialfall der Sigmoid-Funktion:

$$\text{act}(t) = \frac{1}{1 + e^{-t}}. \quad (3.12)$$

Die Vorteile der logistischen Funktion sind, dass der Ausgangswert stets zwischen 0 und 1 liegt und dass keine Sprünge im Gradienten auftreten. Nachteilig ist, dass der Gradient für sehr große oder sehr kleine Eingangswerte verschwindet (engl. *vanishing gradient problem* [KK01]). Dies kann zu Problemen wie z. B. einer langsamen Trainingskonvergenz führen. Des Weiteren ist die logistische Funktion verglichen mit anderen Aktivierungsfunktionen rechenaufwendig.

ReLU und PReLU Die ReLU-Aktivierungsfunktion (engl. *rectified linear unit*) und ihre Erweiterung, die PReLU (engl. *parameterized ReLU*), werden häufig verwendet. Die Formel lautet [He+15]:

$$\text{act}(t) = \max(0, t) + \alpha \cdot \min(0, t), \quad (3.13)$$

wobei für die ReLU $\alpha = 0$ und für die PReLU $\alpha \neq 0$ gilt. Der Rechenaufwand der ReLU ist geringer, der Gradient verschwindet jedoch vollständig im Falle negativer Eingangswerte. Der Trainingsprozess stagniert dann. Dieses Problem wird mit der Erweiterung der ReLU zur PReLU behoben. Der Parameter α wird während des Trainingsprozesses ebenfalls optimiert.

3.2.2 Faltende neuronale Netze

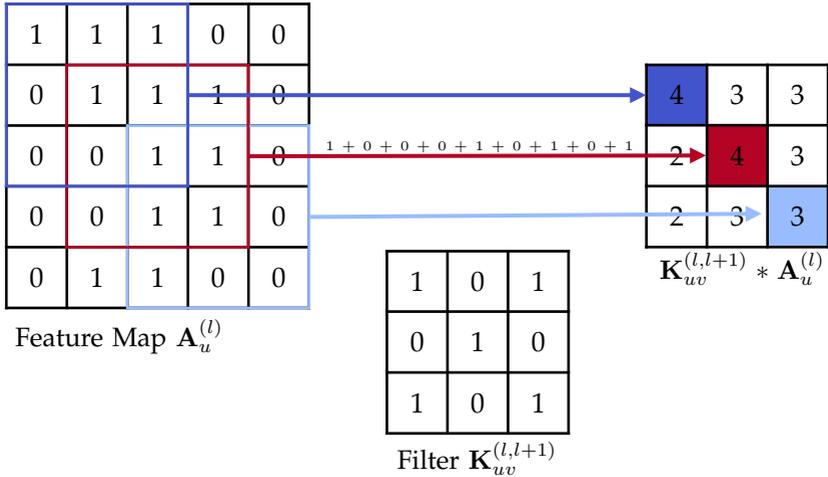
Faltende neuronale Netze sind eine spezielle Art künstlicher neuronaler Netze, die besonders gut für die Verarbeitung von Informationen in Bildern geeignet sind. Sie zeichnen sich durch die folgenden vier Eigenschaften aus [LBH15]: lokale Verbindungen, geteilte Gewichte, Pooling und Verwendung vieler Schichten. In den letzten Jahren wurden CNNs mit großem Erfolg für die Detektion, Segmentierung und Erkennung von Objekten in Bildern eingesetzt. Berühmte Architekturen sind beispielsweise LeNet [LeC+98], AlexNet [KSH17] und das VGG-Netz [SZ15]. Auf die Grundelemente eines CNNs, die Faltungs- und die Pooling-Schichten, wird nun detaillierter eingegangen. Die Grundideen der beiden Operationen sind in Abbildung 3.4 veranschaulicht.

3.2.2.1 Faltungsschichten

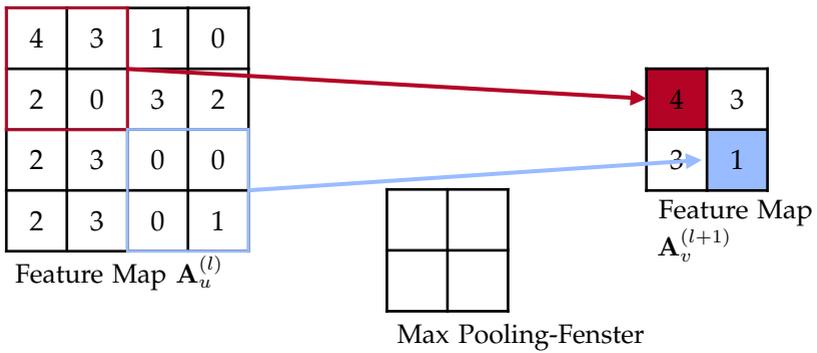
Am Eingang des CNNs wird ein Tensor mit den Dimensionen (Bildbreite) \times (Bildhöhe) \times (Kanäle) übergeben. Abweichend von der Notation in der restlichen Arbeit werden die Neuronen der Eingangs- und aller weiteren Faltungs- und Pooling-Schichten hier nicht als Vektor, sondern als Tensor angeordnet. Die verdeckten Schichten enthalten jeweils mehrere Tensoren zweiter Stufe, die sogenannten Feature Maps (dt. Merkmalskarten). Die Feature Maps einer Schicht sind alle gleich groß, ihre Anzahl wird auch als Tiefe der Schicht bezeichnet. Abweichend von den vollständig verbundenen Netzen wird die Aktivierung der Neuronen einer Feature Map $\mathbf{A}_v^{(l+1)}$ der Schicht $l + 1$ berechnet, indem ein Filter über die gesamte räumliche Dimension der Vorgängerschicht geschoben und anschließend über deren Tiefe addiert wird [LKF10]:

$$\mathbf{A}_v^{(l+1)} = \text{act} \left(\mathbf{B}_v^{(l+1)} + \sum_u \mathbf{K}_{uv}^{(l,l+1)} * \mathbf{A}_u^{(l)} \right). \quad (3.14)$$

Zu beachten ist, dass u hier die Feature Maps der Schicht l indiziert und v die Feature Maps der Schicht $l + 1$ und nicht die einzelnen enthaltenen Neuronen. Jede Feature Map \mathbf{A}_u der Schicht l wird mit einem Faltungskern $\mathbf{K}_{uv}^{(l,l+1)}$ (dem Filter) gefaltet. Die Neuronen aufeinander folgender Schichten eines CNNs sind also durch die Parameter dieser Faltungskerne miteinander verbunden. Sie müssen während des Trainings optimiert



(a) Faltung mit einem Filter der Größe 3×3 und der Schrittweite 1.



(b) Max Pooling mit einer Fenstergröße von 2×2 und einer Schrittweite von 2.

Abbildung 3.4 Beispiele für eine Faltung und für Max Pooling.

werden. Die Ergebnisse der Faltung werden über alle Feature Maps der Schicht l aufsummiert. Nachdem zusätzlich der Bias $\mathbf{B}_v^{(l+1)}$ addiert und die nicht lineare Aktivierungsfunktion (auf jedes Neuron) angewendet wurde, ergibt sich die Feature Map \mathbf{A}_v der Schicht $l + 1$.

Da die Faltungskerne „kleiner“ sind als die Feature Maps, häufig verwendete Größen sind 3×3 oder 7×7 , reagiert jedes Neuron der Schicht $l + 1$ nur auf einen kleinen Bereich der Feature Map der vorausgehenden Schicht. Deshalb spricht man von einem lokalen Sichtfeld oder lokalen Verbindungen (engl. *local receptive field*). Es sind also nicht alle Neuronen einer Schicht mit allen anderen verbunden. Benachbarte Neuronen innerhalb einer Feature Map werden aus überlappenden Bereichen der Vorgängerschicht berechnet. Alle Neuronen einer Feature Map teilen sich dieselben Parameter (denselben Filter). Das ist bei vollständig verbundenen Netzen anders. Die kleinen Filterkerne und die gemeinsam genutzten Gewichte haben zur Folge, dass pro Schicht weniger Parameter trainiert werden müssen und deshalb tiefere Netzarchitekturen, also mehr Schichten, möglich sind als bei vollständig verbundenen Architekturen [HJ17].

3.2.2.2 Pooling-Schichten

Pooling-Schichten werden oft im Wechsel mit Faltungsschichten eingesetzt. Beim Pooling (dt. Bündelung, Zusammenfassung) wird jede Feature Map separat behandelt. Die Anzahl bleibt also unverändert. Ziel des Poolings ist es, die Größe einer Feature Map zu verringern. Dies wird durch das Zusammenfassen benachbarter Neuronen erreicht. Die Größe des Fensters, innerhalb dessen die Neuronen zusammengefasst werden, kann frei gewählt werden. Üblich ist eine Fenstergröße von 2×2 und eine der Fenstergröße entsprechenden Schrittweite, mit der das Pooling-Fenster über die Eingangsdaten „geschoben“ wird. Es gibt verschiedene Möglichkeiten, die Neuronen innerhalb des Fensters zusammenzufassen: Beim Average Pooling wird die durchschnittliche Aktivierung verwendet, beim Max Pooling wird die größte vorhandene Aktivierung übernommen. Letzteres hat sich gegenüber anderen Pooling-Methoden als vorteilhaft erwiesen [SMB10]. Eine weitere Möglichkeit ist es, keine separaten Pooling-Schichten zu verwenden, sondern stattdessen ebenfalls Faltungsschichten einzusetzen. In diesem Fall verwendet man Filter größerer Schrittweite, damit die Faltung

nicht auf jedes, sondern z. B. bei einer Schrittweite von 2 nur auf jedes zweite Neuron angewendet wird [Spr+15].

Die Verwendung von Pooling hat mehrere Gründe. Zum einen wird das neuronale Netz so robuster gegenüber kleinen Positionsänderungen der Merkmale innerhalb des Eingangsbildes [LKF10]. Außerdem reduziert sich der Rechenaufwand, weil die Anzahl der Neuronen verringert wird. Zusätzlich wird durch Pooling das lokale Sichtfeld tieferer Schichten vergrößert.

Mit einem CNN können komplexe Merkmale des Eingangsbildes durch die Kombination einfacher Merkmale extrahiert werden. In der ersten Schicht werden nur einfache Merkmale, z. B. Kanten, detektiert. In der darauffolgenden Schicht werden diese Merkmalsextraktoren zusammengesetzt und dadurch werden komplexere Strukturen detektierbar [Arr+16]. Außerdem vergrößert sich der für die Neuronen sichtbare Bildbereich mit jeder Faltung und jeder Pooling-Operation. Je tiefer eine Schicht also im neuronalen Netz liegt, desto komplexer sind die dort erkennbaren Muster und desto größer ist das Sichtfeld der Neuronen. Für weitere Informationen über CNNs sei auf die Arbeit von LeCun et al. [LBH15] verwiesen.

3.2.3 Implementierungsdetails

Im Folgenden wird auf Details und Begriffe eingegangen, die zum Verständnis der in Kapitel 5 beschriebenen CNN-Architekturen beitragen.

Zero Padding Um bei der Faltung eine unerwünschte Verkleinerung der räumlichen Dimension zu vermeiden, wie sie in Abbildung 3.4(a) auftritt, wird sogenanntes Zero Padding eingesetzt. Der Rand des Eingangstensors wird so mit Nullen aufgefüllt, dass die Größe vor und nach der Faltung dieselbe ist [DV18].

Vektorisierung, Global Average Pooling Auf die Faltungs- und Pooling-Schichten, die Merkmale aus den Eingangsdaten extrahieren, folgen normalerweise mehrere vollständig verbundene Schichten. Diese ordnen die extrahierten Merkmale verschiedenen Ausgängen zu. Lange war es üblich, die Feature Maps der letzten Faltungsschicht dafür zu vektorisieren (engl.

flattening) und anschließend an die vollständig verbundenen Schichten zu übergeben. Ein großer Nachteil dieses Ansatzes ist die enorme zusätzliche Anzahl zu trainierender Modellparameter und das damit verbundene Risiko der Überanpassung. Dieses Problem kann mit Global Average Pooling (GAP) umgangen werden: Anstatt alle Neuronen aller Feature Maps zu erhalten, wird für jede Feature Map nur die durchschnittliche Aktivierung berechnet. Die gesamte Feature Map wird also auf ein einziges Neuron reduziert. Dieses kann anschließend direkt einer Kategorie zugeordnet oder an vollständig verbundene Schichten übergeben werden. Zusätzlich zur Reduktion des Überanpassungsrisikos wird das Modell durch GAP robuster gegenüber räumlichen Translationen des Eingangsbildes [LCY14].

Gradientenverfahren und Rückwärts-Propagation Beim Training neuronaler Netze geht es, analog zu allen anderen Modellen des überwachten maschinellen Lernens, darum, die Differenz zwischen dem Label und dem Output des Modells zu minimieren. Dafür muss zunächst eine Verlustfunktion E definiert werden (siehe Abschnitt 3.1.2). Diese wird während des Modelltrainings durch die Optimierung der Modellparameter minimiert. Als Optimierungsverfahren wird bei neuronalen Netzen meist ein stochastisches Gradientenverfahren (engl. *stochastic gradient descent*, SGD) eingesetzt [LKF10].

Bei Gradientenverfahren wird der Einfluss kleiner Modellparameteränderungen auf die Verlustfunktion betrachtet. Dieser Einfluss wird durch den Gradienten der Verlustfunktion in Abhängigkeit der Modellparameter \mathbf{W} quantifiziert. Im einfachsten Fall werden die Modellparameter iterativ mit einer konstanten Lernrate $\lambda > 0$ angepasst [LeC+98]:

$$\mathbf{W}_k = \mathbf{W}_{k-1} - \lambda \frac{\partial E(\mathbf{W})}{\partial \mathbf{W}}. \quad (3.15)$$

Anstatt den Gradienten basierend auf dem gesamten Trainingsdatensatz zu berechnen, wird er beim SGD-Verfahren oft nur auf Basis einer Teilmenge (einem sogenannten Mini-Batch) berechnet. Normalerweise führt diese Methode zu einer schnelleren Konvergenz [LeC+98]. Gradientenverfahren werden schon seit Jahrzehnten verwendet, waren jedoch zumeist auf lineare Modelle limitiert [DHS00]. Dass dieses einfache Verfahren für das Training komplexer ML-Modelle geeignet ist, hat zwei Gründe [LeC+98]:

Erstens scheint die Anwesenheit lokaler Minima in der Verlustfunktion in der Praxis kein großes Problem zu sein. Zweitens existiert mit der Rückwärts-Propagation (engl. *backward propagation*) des Fehlers eine einfache und recheneffiziente Methode zur Berechnung des Gradienten in nicht linearen, aus mehreren Schichten bestehenden Systemen. Diese wurde 1986 von Rumelhart et al. [RHW86] beschrieben und wird noch heute verwendet. Verschiedene Methoden für ein effektives Modelltraining werden von LeCun et al. [LeC+12] diskutiert. Der in dieser Arbeit verwendete Adam-Optimierer [BK15] ist eine Weiterentwicklung des SGD-Verfahrens.

Batch-Normalisierung Das Training eines neuronalen Netzes wird durch die Tatsache erschwert, dass jede Schicht durch die Parameter aller vorhergehenden Schichten beeinflusst wird. Während des Trainings verändern sich die Parameter und damit auch die Häufigkeitsverteilung der Eingangswerte einer Schicht. Das ist problematisch, weil sich die Schicht dann immer wieder an diese neue Verteilung anpassen muss. Je tiefer ein Netz ist, desto größer ist dieses Problem, weil selbst kleine Änderungen bei der Propagation durch viele Schichten verstärkt werden und zu signifikanten Änderungen in tieferen Schichten führen können. Ioffe und Szegedy [IS15] entwickelten die Batch-Normalisierung (BN) mit dem Ziel, diesen Effekt abzuschwächen. Da das stochastische Gradientenverfahren zumeist auf Teilmengen (den Mini-Batches) angewendet wird, wird hier analog vorgegangen. Der Mittelwert μ_{Batch} und die Standardabweichung σ_{Batch} des Batches werden für jedes Neuron berechnet. Anschließend wird die Aktivierung des Neurons damit normalisiert, mit einem Parameter χ multipliziert und zum Schluss wird ein Parameter ϕ addiert:

$$a_j = \frac{a_i - \mu_{\text{Batch}}}{\sigma_{\text{Batch}}} \chi + \phi. \quad (3.16)$$

Batch-Normalisierung führt zu einer Häufigkeitsverteilung der Eingangswerte mit dem Mittelwert ϕ und der Standardabweichung χ . ϕ und χ zählen zu den Modellparametern und werden während des Trainingsprozesses optimiert. Nach dem Training werden μ_{Batch} und σ_{Batch} auf einen festen Wert gesetzt. Durch die Anwendung von BN-Schichten im Training müssen sich tiefere Schichten weniger an Veränderungen in davor liegenden Schichten anpassen, es kann eine höhere Lernrate gewählt werden, die Trainingszeit verringert sich und die Abhängigkeit des Trainingsprozesses

von der Initialisierung sinkt [IS15]. Batch-Normalisierung wird sehr häufig eingesetzt, auch wenn noch nicht alle zugrunde liegenden Mechanismen verstanden sind [San+19].

Dropout Wird ein tiefes neuronales Netz mit wenigen Beispielen trainiert, kommt es häufig zu einer Überanpassung an die Trainingsdaten und einer daraus resultierenden schlechten Modellgüte auf den Validierungs- und Testdaten. Dropout ist eine Regularisierungsmethode, mit der die Überanpassung verringert werden kann. Dabei werden während des Trainingsprozesses zufällig ausgewählte Neuronen weggelassen, beispielsweise 50 %. Dies führt dazu, dass eher allgemein hilfreiche als sehr spezifische Merkmale erlernt werden und das Modell besser generalisiert [Hin+12].

3.2.4 Highway-Schicht

In einer Highway-Schicht wird die Formel (3.10) erweitert zu [SGS15]:

$$\begin{aligned} \mathbf{a}^{(l+1)} = & H_{l,l+1} \left(\mathbf{a}^{(l)}, \mathbf{W}_{H_{l,l+1}} \right) \circ T_{l,l+1} \left(\mathbf{a}^{(l)}, \mathbf{W}_{T_{l,l+1}} \right) \\ & + \mathbf{a}^{(l)} \circ \left(\mathbf{1} - T_{l,l+1} \left(\mathbf{a}^{(l)}, \mathbf{W}_{T_{l,l+1}} \right) \right). \end{aligned} \quad (3.17)$$

$\mathbf{a}^{(l)}$, $\mathbf{a}^{(l+1)}$, $H_{l,l+1}(\mathbf{a}^{(l)}, \mathbf{W}_{H_{l,l+1}})$ und $T_{l,l+1}(\mathbf{a}^{(l)}, \mathbf{W}_{T_{l,l+1}})$ müssen dieselbe Dimension haben. T wird als Transfergate, $1 - T$ als Carrygate bezeichnet. Multiplikation, Addition und Subtraktion erfolgen elementweise. In Abbildung 3.5 ist anschaulich gezeigt, was die Formel bedeutet. Der Name Highway kommt daher, dass der Ausgang der Schicht gerade dem Eingang entspricht, wenn $T_{l,l+1}(\mathbf{a}^{(l)}, \mathbf{W}_{T_{l,l+1}}) = 0$. Dann entsteht ein Informations-Highway, der die Eingangs-Information unverändert an die nächste Schicht weitergibt. Falls $T_{l,l+1}(\mathbf{a}^{(l)}, \mathbf{W}_{T_{l,l+1}}) = 1$ ist, entspricht die Transformation wieder Formel (3.10). In Abhängigkeit des Transfergates verändert also die ganze Schicht ihr Verhalten. In dieser Arbeit wird ein Autoencoder mit mehreren Highway-Schichten zur Korrektur von Fehlern in den Rautiefenmesswerten verwendet. Dies ist in Abschnitt 4.3.1 detailliert beschrieben.

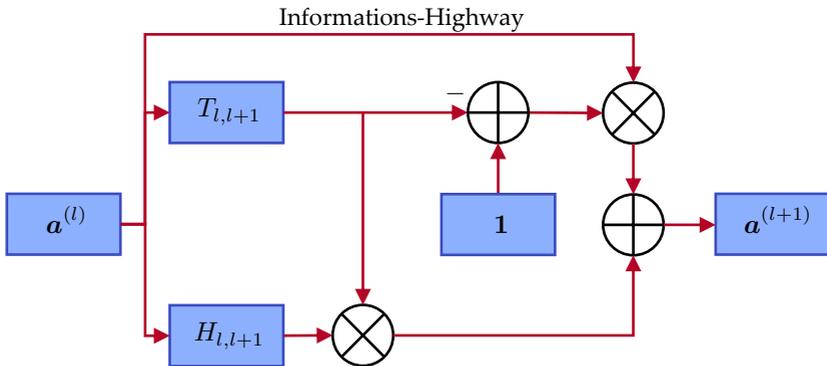


Abbildung 3.5 Architektur einer Highway-Schicht, die ihr Verhalten in Abhängigkeit des Transfergates $T_{l,l+1}$ verändert.

3.2.5 Anwendungen neuronaler Netze in der Fertigung

Fertigungsprozesse sind in den letzten Jahren automatisierter, computerisierter und komplexer geworden. Es entsteht die neue Form der intelligenten Fertigung, welche eine weitere industrielle Revolution einleitet [Kus18]. Dabei kommen zunehmend auch datenbasierte Methoden des maschinellen Lernens zum Einsatz. Kusiak [Kus18] schrieb 2018: „We are witnessing the renaissance of data in manufacturing. (...) It will be the best source for preserving and extraction of past and new knowledge related to manufacturing.“

Methoden des maschinellen Lernens werden bereits in verschiedensten Bereichen der Fertigung eingesetzt, unter anderem zur Prozessoptimierung, -überwachung und -regelung und zur vorausschauenden Wartung (engl. *predictive maintenance*) [Wue+16]. Je nach Problemstellung und den zur Verfügung stehenden Daten werden verschiedene Modellansätze verwendet. Die in Abschnitt 3.2.2 eingeführten faltenden neuronalen Netze finden breite Anwendung bei der Prozessüberwachung. Sie eignen sich hervorragend zur Detektion, Lokalisierung und Klassifizierung von Defekten in verschiedensten Produktionsprozessen. Im Gegensatz zu anderen Detektionsmethoden, die auf manuell designten Merkmalen beruhen, lernen die Netze während des Trainings, relevante Merkmale zu extrahieren und den Defekt darauf basierend zu klassifizieren.

3.2.5.1 Anwendungsbeispiele von CNNs im Fertigungskontext

CNNs wurden bereits eingesetzt, um LED-Chips auf zwei Defektarten (Linienfehler, Kratzer) hin zu untersuchen [Lin+19]. Als Eingangsinformation für das CNN diente ein Graustufenbild des Chips. Es konnte eine Klassifikationsgenauigkeit von 95 % erzielt werden. Nur in wenigen Arbeiten wird über die Detektion und Klassifikation hinaus versucht, die Ursache für den Defekt zu finden. Ein Beispiel dafür ist die Arbeit von Lee et al. [LCK17], die den Einsatz eines CNNs für die Fehlerklassifizierung und -diagnose in der Halbleiterherstellung thematisiert. Dem CNN wurden keine Bilder präsentiert, sondern eine zweidimensionale Matrix verschiedener zeitabhängiger Sensorsignale. Auf einer Achse war die Zeit, auf der anderen waren die unterschiedlichen Sensorsignale aufgetragen. Das Sichtfeld des CNNs bewegte sich nur entlang der Zeitachse. Zusätzlich zur Detektion eines Fehlers wurde das für die Entscheidung des CNNs wichtigste Sensorsignal und damit indirekt die Fehlerursache identifiziert. Dies geschah durch die Betrachtung der Gewichte zwischen der Eingangs- und der ersten verdeckten Schicht in Kombination mit den Aktivierungen. Der verwendete Datensatz beinhaltete 6000 fehlerfreie und gleich viele fehlerbehaftete Beispiele. Die Modellgüte unterschied sich für verschiedene Fehlertypen, im Mittel konnte eine Klassifikationsgenauigkeit von 98 % erreicht werden. Neben der Erkennung von Defekten werden neuronale Netze auch vermehrt zur kontaktlosen Bestimmung von Oberflächenkenngrößen, z. B. der Rautiefe, eingesetzt. Sun et al. [Sun+18] bestimmten die Rauheit von gefrästen Oberflächen mithilfe eines CNNs nach einer aufwendigen Schräglagenkorrektur und Filterung der aufgenommenen Bilder. Die für das Training des Modells benötigten Labels wurden durch eine Tastspitzenmessung ermittelt. Der Datensatz bestand aus 2040 Bildern, die mittlere Korrektorklassifikationsrate betrug 95 %.

Andere Anwendungsbereiche von CNNs in der Fertigung sind die Prädiktion der verbleibenden Nutzungsdauer, z. B. von Turbofantriebwerken [SZL16] und die Zustandsüberwachung, z. B. von Getriebezähnen [KC19]. In beiden Fällen wurden Sensorsignale als Eingangsinformation für die neuronalen Netze verwendet. Bei Sateesh Babu et al. [SZL16] waren es die Zeitreihen von 21 Sensoren, die von Wang et al. [WYH12] näher beschrieben wurden. [KC19] betrachten zwei Zeitreihen-Signaltypen, Vibrationen und den sogenannten Übertragungsfehler des jeweiligen Getriebes.

3.2.5.2 Anwendungsbeispiele von CNNs im Kontext der Lasermaterialbearbeitung

Im Kontext der Lasermaterialbearbeitung werden CNNs bislang vor allem zur Überwachung des Schweißprozesses eingesetzt. Ein Bild der Schweißnaht wird an das Netz übergeben und hinsichtlich verschiedener Qualitätskriterien klassifiziert. Bei Khumaidi et al. [KYP17] waren das die Klassen: gute Qualität, verspritzt, porös und Hinterschnitt. Sizyakin et al. [Siz+19] betrachteten Risse, Poren, Einschlüsse, eine fehlende Verschweißung und Verformungen. Während Khumaidi et al. [KYP17] eine experimentelle Datenbasis bestehend aus 120 Schweißnähten für ihr Modelltraining und die Evaluierung verwendeten, erfolgte das Training bei Sizyakin et al. [Siz+19] auf der öffentlich verfügbaren Datenbasis GDXray [Mer+15], die 88 Röntgenbilder von Schweißverbindungen enthält. Eine Online-Methode zur Defektdetektion mit CNNs beim Schweißen wurde von Zhang et al. [ZWC19] vorgeschlagen.

Stahl und Jauch [SJ19] veröffentlichten eine Arbeit zur bildbasierten Bestimmung der Rauheit einer Laserschnittkante mit einem CNN, ähnlich dem in Abschnitt 6.2.2 beschriebenen Ansatz. Sie übergaben ein Bild einer Laserschnittkante an ein CNN und schätzten die Rauheit. Ihr Ansatz hat jedoch die folgenden Schwächen: Zum einen wurde die gesamte Kante durch einen gemittelten Rauheitswert beschrieben, welcher wenig aussagekräftig ist, da Laserschnittkanten oft mehrere Schichten verschiedener Rauheit aufweisen. Außerdem haben die Autoren einen Fehler bei der Aufteilung in Trainings- und Testdaten gemacht: Zu jeder Schnittkante gibt es fünf Bilder, die sehr ähnlich aussehen, da die gesamte Kante mit denselben Prozessparametern geschnitten wurde. Die Anzahl der Bilder wurde außerdem durch Spiegeln an der vertikalen Achse weiter erhöht. Anschließend wurden die Bilder zufällig den Trainings- oder Testdaten zugeordnet, weshalb davon auszugehen ist, dass die Schätzung der Rauheit durch das CNN nur deshalb so gut gelang, weil es die Schnittkanten aus dem Training im Test wiedererkannt hat. Beim dem von Stahl und Jauch [SJ19] verwendeten CNN handelte es sich um ein modifiziertes AlexNet [KSH17], welches viel mehr Parameter aufweist als das in dieser Arbeit verwendete Modell. Die Datenbasis bestand aus 100 Teilen.

Eine ganz andere und bislang wenig eingesetzte Anwendung stellt die Prädiktion des Prozessresultats mithilfe eines CNNs dar, wie sie von

Gonçalves et al. [GSP20] publiziert wurde. Die Autoren schätzten die finale Schmelzperlengeometrie beim Laserstrahl-Auftragschweißen (engl. *laser cladding*). Bei dieser Fertigungstechnik wird ein Volumenaufbau durch ein Schweißzusatzpulver realisiert. Als Eingangsinformation für das CNN diente ein Bild des Schmelzbades, das koaxial zum Laserstrahl aufgenommen wurde. Es wurde ergänzt durch verschiedene Prozessparameter. Am Ausgang des Netzes wurden die Breite und Höhe der zu erwartenden Schmelzperle geschätzt. Gonçalves et al. [GSP20] verglichen sechs CNNs, die sich hinsichtlich ihrer Anzahl der Faltungs- und der vollständig verbundenen Schichten unterschieden. Am besten schnitt ein Netz mit fünf Faltungs- und einer vollständig verbundenen Schicht ab. Als Datengrundlage dienten 72 Schmelzperlen.

3.2.6 Interpretierbarkeit neuronaler Netze

Dank der zunehmenden Verfügbarkeit großer Datenmengen, Verbesserungen der Methodik und immer mehr Rechenkapazität konnten mit komplexen ML-Modellen (vor allem mit künstlichen neuronalen Netzen) in verschiedensten Bereichen Ergebnisse erzielt werden, die bestehende Lösungen und/oder die Leistungsfähigkeit des Menschen übertreffen. Ein Problem ist jedoch, dass neuronale Netze aufgrund ihrer verschachtelten, nicht linearen Struktur als sogenannte Black Boxes gelten, deren inneres Verhalten unbekannt ist. Es gibt viele Gründe, warum diese mangelnde Transparenz problematisch und die Interpretierbarkeit komplexer Modelle wünschenswert oder sogar notwendig ist.

Zum einen lässt sich so erkennen, wenn ein neuronales Netz „falsche“ Zusammenhänge erlernt hat, die jedoch zu richtigen Klassifikationen führen. Beispielsweise wurden Pferde im PASCAL VOC 2007 Datensatz [Eve+10] sehr zuverlässig erkannt, was jedoch an den Copyright-Hinweisen lag, die nur diese Bilder enthielten [Lap+19]. Sobald diese entfernt wurden, konnten Pferde nicht mehr korrekt klassifiziert werden.

Des Weiteren sind Visualisierungen und Erklärungen des Verhaltens neuronaler Netze wichtig, um das Vertrauen in die Technologie zu stärken. Die Erklärung von Entscheidungen ist ein zentraler Bestandteil zwischenmenschlicher Interaktion und sollte deshalb auch für Mensch-Maschine-Interaktionen gelten [Ant+12]. Dies betrifft insbesondere Systeme, die

Menschen unterstützen sollen und sicherheitsrelevante Anwendungen wie das autonome Fahren.

Die Erklärbarkeit von Modellentscheidungen ist außerdem wichtig für die Klärung rechtlicher und ethischer Fragen, beispielsweise nach der Haftung oder nach diskriminierenden Elementen [SM19].

Der letzte und im Kontext dieser Arbeit wichtigste Grund, Transparenz zu fordern, ist, dass Einblicke in die Modelle die Grundvoraussetzung für neue Erkenntnisse sind. KI-basierte Modelle sind in der Lage, Muster, Strukturen und Regeln in Daten zu erkennen, die bisher gar nicht bekannt waren. Samek und Müller [SM19] schreiben, dass die so erlangten neuen Erkenntnisse in der Wissenschaft oft wichtiger sind als die Vorhersage selbst. In der Vergangenheit wurden deshalb oft einfache, leicht interpretierbare Modelle eingesetzt. Neue Möglichkeiten, komplexe Modelle zu visualisieren, eröffnen ganz neue Perspektiven.

In den vergangenen Jahren wurde aus den oben genannten Gründen vermehrt an Ansätzen zur Erklärbarkeit und Visualisierung künstlicher neuronaler Netze geforscht. Für einen ersten Überblick ist es sinnvoll, bestehende Ansätze, wie von Samek und Müller [SM19] vorgeschlagen, nach ihrem Informationsgehalt und nach der zugrunde liegenden Methodik zu unterscheiden. Es handelt sich nicht um einen vollständigen Überblick, sondern um eine kurze Zusammenstellung wichtiger Methoden.

3.2.6.1 Visualisierungsmethoden nach ihrem Informationsgehalt

Erklärung der erlernten Repräsentationen Bei diesem Ansatz geht es darum, die von einem trainierten Netz erlernten Muster zu verstehen. Dabei werden zum Beispiel einzelne Neuronen oder Gruppen mehrerer Neuronen betrachtet [Bau+17]. Eine andere Herangehensweise ist die Berechnung eines Prototyps, der für ein bestimmtes erlerntes Konzept repräsentativ ist [SVZ14].

Erklärung einzelner Vorhersagen Durch die Visualisierung der für die jeweilige Entscheidung des Netzes wichtigsten Pixel ausgewählter Eingangsbilder in einer Heatmap kann ein neuronales Netz validiert werden [MSM18]. Dadurch kann außerdem das Vertrauen in seine Funktionalität erhöht werden und es können neue Muster oder Strategien identifiziert und neue wissenschaftliche Erkenntnisse gewonnen werden. Zu diesen

Techniken zählt unter anderem die in dieser Arbeit eingesetzte Layer-wise Relevance Propagation (siehe Abschnitt 5.2).

Erklärung des Modellverhaltens Dieser Ansatz geht über die Erklärung einzelner Vorhersagen hinaus und versucht, das Modellverhalten, beispielsweise die angewandten Strategien, besser zu verstehen. Lapuschkin et al. [Lap+19] generieren diese Information durch das Clustern mehrerer einzelner Vorhersagen und der zugehörigen Heatmaps.

Erklärung mithilfe repräsentativer Beispiele Hierbei geht es darum, den Trainingsdatensatz und dessen Einfluss auf das neuronale Netz besser zu verstehen. Dazu werden repräsentative Trainingsbeispiele identifiziert und analysiert [KL17].

3.2.6.2 Visualisierungsmethoden nach der zugrunde liegenden Methodik

Stellvertreter-Modelle (engl. *surrogates*) Da einfache (z. B. lineare) Modelle leicht zu interpretieren sind, werden komplexe Modelle lokal mit einer einfachen Funktion approximiert. Ein bekanntes Beispiel dafür ist die LIME-Methode (engl. *local interpretable model-agnostic explanations*) [RSG16].

Lokale Störungen (engl. *local perturbations*) Ein Modell wird erklärt, indem seine Reaktionen auf lokale Änderungen untersucht werden. Es ist jedoch zu beachten, dass damit weniger die Vorhersage selbst, sondern nur Änderungen der Vorhersage erklärt werden können. Das berühmteste und älteste Beispiel ist die Sensitivitätsanalyse [Mor+95]. Diese wird aufgrund verschiedenster Probleme als ungeeignet für die Erklärung heutiger ML-Modelle angesehen [MSM18]. Verwandt damit sind Methoden, die das Modellverhalten in Abhängigkeit allgemeinerer Störungen untersuchen. Bei der Okklusionsmethode [ZF14] wird beispielsweise ein Teil des Eingangsbildes abgedeckt. Die oben bereits erwähnten Prototyp-Methoden fallen ebenfalls in diese Kategorie. Der Prototyp einer Klasse ist der Input, der die maximale Antwort des Modells bei dieser Klasse hervorruft. Die

Erklärung des Modells wird in diesem Fall als Optimierungsproblem betrachtet. Prototypen für Kategorien von Alltagsobjekten, denen mithilfe eines CNNs Bilder zugeordnet wurden, zeigten Simonyan et al. [SVZ14].

Propagationsbasiert Propagationsbasierte Ansätze integrieren die Modellstruktur in den Erklärungsprozess. Bei der in dieser Arbeit verwendeten LRP-Technik wird die Modellstruktur zusammen mit geeigneten Propagationsregeln verwendet, um zu berechnen, wie groß der Einfluss jedes Eingangsmerkmals auf die Modellvorhersage war. Andere Methoden dieser Kategorie sind Deconvolution [ZF14] und Guided Backpropagation [Spr+15]. Mittlerweile sind viele dieser Methoden mithilfe der Python-Bibliothek iNNvestigate [Alb+19] anwendbar. Zum Zeitpunkt der Implementierung der LRP im Rahmen dieser Arbeit war das noch nicht der Fall.

Meta-Methoden Um allgemeine Strategien und Muster zu identifizieren, werden einzelne Erklärungen zusammengefasst und gemeinsam analysiert. Ein bekannter Vertreter der Meta-Methoden ist die SpRAY (engl. *spectral relevance analysis*) [Lap+19]).

3.2.6.3 Anwendungen der Layer-wise Relevance Propagation

Bei der Visualisierung neuronaler Netze gibt es nicht die eine immer passende Methode. Stattdessen muss je nach Fragestellung, Modellarchitektur und Datengrundlage entschieden werden, welche Herangehensweise am besten geeignet ist. In dieser Arbeit wurden verschiedene Visualisierungsmethoden getestet. Neben der LRP waren das Regression Activation Maps [Zho+16] und Class Activation Maximization [SVZ14]. Letztere lieferten jedoch keine für den Menschen verständlichen Ergebnisse. Die Gründe dafür werden in Abschnitt 6.4.1 kurz thematisiert.

Neben den Publikationen, in denen die Grundlagen der LRP erläutert und veranschaulicht werden, gibt es auch Arbeiten, die diese Visualisierungstechnik bereits in der Praxis anwenden. Eine davon wurde von Grezmač et al. [Gre+20] publiziert. Als Eingangsdaten für das CNN dienten die in ein Bild überführten Vibrationssignale eines Asynchronmotors. Das CNN klassifizierte, ob der Motor in Ordnung war oder ob eines von drei

Problemen vorlag. Mithilfe der LRP konnte identifiziert werden, welcher Signalanteil für welchen Fehlertyp besonders relevant war. Außerdem konnte untersucht werden, inwieweit sich die relevanten Bereiche verschiedener, aber zum selben Fehlertyp gehörender Eingangsbilder unterschieden und ob es Unterschiede bei verschiedenen CNN-Architekturen gab. Ein Beispiel aus einem anderen Bereich, der Medizin, wurde von Böhle et al. [Böh+19] veröffentlicht. Sie setzten die LRP ein, um die Entscheidungen eines CNNs zu erklären, das Anzeichen für Alzheimer in MRT-Bildern des Gehirns erkennt. Die Hirnareale, die in den Bildern anhand der LRP als wichtig für die Entscheidung des CNNs identifiziert werden konnten, waren in anderen Arbeiten bereits als Hinweise auf Alzheimer beschrieben worden.

In beiden Fällen wurde die Visualisierung genutzt, um neue Erkenntnisse zu erzielen. Im ersten Beispiel ging es darum, diese Erkenntnisse für die Optimierung der Modelle zu nutzen. Im zweiten Beispiel unterstützte die LRP den Menschen bei einer Entscheidung. Ein Arzt oder eine Ärztin kann von einem neuronalen Netz, das Alzheimer erkennt, lernen, welche Bildbereiche besonders wichtig sind. Des Weiteren stärkt die Visualisierung das Vertrauen in ML-Modelle, weil die Entscheidung des Netzes begründet wird. In dieser Arbeit wird die LRP genutzt, um das Prozessverständnis zu erweitern und um bei Problemen zu unterstützen (siehe Abschnitt 6.4.1).

4 Generierung der Datenbasis

Als Grundlage für das Training der ML-Modelle wurde eine große Datenbasis erzeugt. Diese sollte möglichst viele unterschiedlich aussehende Schnittkanten enthalten und Wechselwirkungen der Prozessparameter abbilden. Deshalb wurden vier dominante Prozessparameter vollfaktoriell variiert und insgesamt 1126 Teile geschnitten, von denen bei 227 Teilen Fehlschnitte auftraten. Die anderen 899 Teile konnten aus dem Restgitter entnommen und weiter verwendet werden. Von jeder Kante dieser 899 Teile wurden RGB-Bilder und 3D-Punktwolken aufgenommen. Aus den 3D-Punktwolken konnten objektive Qualitätskriterien abgeleitet werden. Außerdem beurteilten fünf Experten die Qualität einiger Kanten subjektiv.

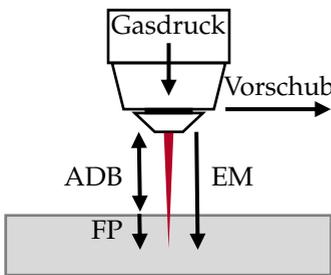
4.1 Schneidexperiment

Um möglichst viele verschiedene Schnittkanten zu erhalten, wurden vier Prozessparameter variiert. Ausgangspunkt für die Parametervariation war der Standardschneidprozess für Edelstahlbleche der Dicke 3 mm auf einer TruLaser 5030 fiber (L68) Maschine. Diese ist mit einem Scheibenlaser der Wellenlänge 1030 nm ausgestattet, einem TruDisk 12001. Weitere Informationen zur Maschine und zum Laser sind in Anhang A.1 zu finden.

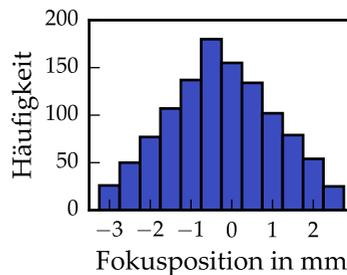
Von TRUMPF werden die in Tabelle 4.1 in der LTT-Spalte aufgelisteten Prozessparameter empfohlen. Ausgehend von den LTT-Werten wurden folgende Prozessparameter variiert: Abstand Düse-Blech (ADB), Einstellmaß (EM), Gasdruck und Vorschub. Die Stufen der Parametervariation sind in Tabelle 4.1 aufgelistet. Die Prozessparameter sind in Abbildung 4.1(a) veranschaulicht. Beim ADB handelt es sich um den Abstand der Düsenspitze von der Blechoberseite. Das Einstellmaß beschreibt den Abstand zwischen Düsenspitze und Strahlfokus. Befindet sich der Fokus unterhalb der Düse, ist das EM negativ, andernfalls hat es ein positives Vorzeichen.

Tabelle 4.1 LTT-Werte und Faktorstufen der wichtigsten Prozessparameter.

Parameter	Einheit	Min.	LTT	Max.	Schritt
Abstand Düse-Blech	mm	0,5	1,5	3,0	0,5
Einstellmaß	mm	-3,5	-2,0	0,5	0,5
Gasdruck	bar	9	15	21	3
Vorschub	m/min	13	21	29	4
Fokusbereich	µm		150		
Laserleistung	kW		8		



(a) Variierte Prozessparameter und FP.



(b) Histogramm der Fokusposition.

Abbildung 4.1 Die variierten Prozessparameter Abstand Düse-Blech (ADB), Einstellmaß (EM), Gasdruck und Vorschub werden durch die Fokusposition (FP), die Kombination von ADB und EM, ergänzt.

Der Gasdruck ist der Druck des als Schneidgas verwendeten Stickstoffs. Der Schneidkopf bewegt sich mit der Vorschubgeschwindigkeit über das Blech, welches ruhig auf den Stegen der Palette aufliegt.

Andere Parameter wurden vor allem aus praktischen Gründen nicht variiert. Die Stahlquelle wurde als gegeben angenommen und zugehörige Parameter, wie z. B. die Stahlqualität und die Laserleistung, wurden nicht verändert. Eine nicht konstante Laserleistung hätte die Kompensation thermischer Effekte in den Optiken (siehe Abschnitt 4.1.3) noch schwieriger gemacht, da deren Erwärmung von der Laserleistung abhängt. Aus demselben Grund wurde die sogenannte Analogwertsteuerung deaktiviert. Normalerweise reduziert diese die Laserleistung, sobald der Vorschub (z. B. an den Ecken) sinkt und verhindert so eine Veränderung der Schnitt-

qualität und die Nichteinhaltung von Formtoleranzgrenzen. Da in dieser Arbeit nur der gerade Schnitt betrachtet wird, spielt dieser Effekt sowieso keine Rolle. Die beweglichen Optiken im Schneidkopf der TruLaser 5030 fiber Maschine hätten eine kleinschrittige Änderung des Fokusdurchmessers ermöglicht. Da dies in vielen herkömmlichen Maschinen nicht der Fall ist, wurde aus Gründen der Übertragbarkeit darauf verzichtet.

Aufgrund der vollfaktoriellen Variation hätte jeder weitere Parameter eine Vervielfachung der Parameterkombinationen bedeutet. Daher musste die Zahl der Parameter begrenzt, die Bandbreite der Variation beschränkt und die Schrittweite möglichst groß gewählt werden. Um sinnvolle Bandbreiten zu wählen, wurden Vorversuche durchgeführt. Hierbei wurde jeweils einer der Parameter ausgehend von der LTT-Parameterkombination variiert. Bei einem Einstellmaß von $-3,8$ mm kam es zu einem Fehlschnitt. Dasselbe passierte bei einem Vorschub von 30 m/min. Es wurde davon ausgegangen, dass das Fehlschnittisiko durch die gleichzeitige Variation weiterer Parameter eher zunimmt. Deshalb wurde ein EM von $-3,5$ mm als Untergrenze und bezüglich des LTT-Werts symmetrisch dazu ein EM von $-0,5$ mm als Obergrenze gewählt. Analog folgte für den Vorschub eine Bandbreite von 13 bis 29 m/min. Durch eine Variation des ADBs und des Gasdrucks konnte kein Fehlschnitt herbeigeführt werden, ohne dass gleichzeitig eine Gefährdung des Schneidkopfs durch Kollisionen oder Verschmutzungen ausgeschlossen werden konnte. Um Kollisionen zu vermeiden, darf die Düse nicht zu nah über die Blechoberfläche bewegt werden. Deshalb wurde ein ADB von $0,5$ mm als Untergrenze festgelegt. Der Gasdruck, der die Schmelze aus dem Schnittspalt austreibt, muss gleichzeitig die Optiken vor Spritzern schützen und darf deshalb nicht zu weit reduziert werden. Dies ist insbesondere wichtig, weil während des Schneidexperiments das stabile Prozessfenster verlassen wird und es bei Fehlschnitten zu starkem Spritzen kommen kann. Darum wurde, wie von erfahrenen Personen empfohlen, kein Gasdruck unter 9 bar benutzt.

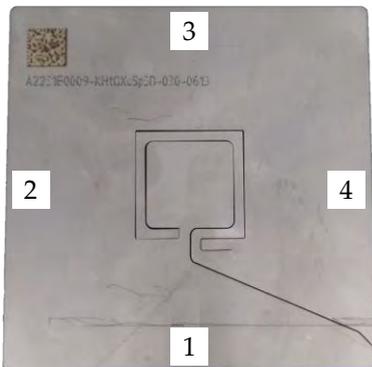
Oft ist es sinnvoll, die oben beschriebenen vier Parameter durch die Fokusposition (FP) zu ergänzen. Diese beschreibt, wie in Abbildung 4.1(a) gezeigt, die Position des Strahlfokus bezogen auf die Blechoberseite. Sie ergibt sich aus der Summe von ADB und EM (unter Beachtung der Vorzeichen). Befindet sich der Strahlfokus inner- oder unterhalb des Blechs, ist die Fokusposition negativ, andernfalls ist sie positiv. Da die vier Pro-

zessparameter ADB, EM, Gasdruck und Vorschub vollfaktoriell variiert wurden, liegt zu jeder Faktorstufe eines Parameters jede Kombination der anderen Parameter vor. Alle Faktorstufen eines Parameters treten gleich häufig auf. Dies trifft auf die Fokusposition nicht zu. Ihr Histogramm ist in Abbildung 4.1(b) dargestellt. Dass auch die Faktorstufen der anderen Parameter später in den Daten nicht mehr gleichverteilt sind, liegt an der Parameterabhängigkeit der Fehlschnitte (siehe dazu Abschnitt 6.1.1). Bei einem Fehlschnitt kann das Blech nicht getrennt und das Teil nicht aus dem Restgitter entnommen werden.

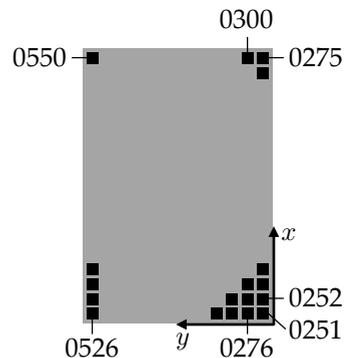
Die vollfaktorielle Variation der vier Prozessparameter (mit sechs, sieben, fünf und fünf Faktorstufen, siehe Tabelle 4.1) ergab 1050 verschiedene Kombinationen. Jede wurde verwendet, um ein Teil zu schneiden. Zusätzlich wurden 5 % zufällig ausgewählte Kombinationen wiederholt. Diese 53 Kombinationen wurden an zufälligen Stellen in den Versuchsplan integriert. Die LTT-Parameterkombination wurde anschließend alle 50 Teile (Teil 0001, 0051, 0101...) eingefügt. Damit konnte schon während des Schneidversuchs überprüft werden, ob sich eine der nicht erfassten Einflussgrößen verändert hatte. RGB-Bilder aller LTT-Teile sind in Anhang A.2 abgebildet. Der für das Schneidexperiment verwendete Versuchsplan enthielt damit insgesamt $1050 + 53 + 23 = 1126$ Parameterkombinationen, von denen 1050 verschieden waren.

4.1.1 Probe

Jede Parameterkombination wurde verwendet, um ein Quadrat mit einer Seitenlänge von 10 cm zu schneiden. Betrachtet werden von jeder Kante nur die mittleren 5 cm, weil nur dort garantiert werden kann, dass alle Prozessparameter konstant waren und der eingestellte Vorschub vorlag. Jedes Teil kann über einen mit dem Laser aufgebracht Data Matrix Code (DMC) identifiziert werden. Dieser codiert die verwendete Maschine, den Versuchsplan, welcher alle Parameterkombinationen enthält, sowie die Blechdicke und die Nummer des Teils. Für alle in dieser Arbeit betrachteten Teile lautet die Identifikationsnummer A2231E0009-KHtQXc5p3D-030-, da sie alle auf derselben Maschine (A2231E0009), mit einem Versuchsplan (KHtQXc5p3D) und in einer Blechdicke (3 mm) hergestellt wurden. Sie unterscheiden sich nur in den letzten vier Ziffern, der Teilnummer von 0001 bis 1126. Deshalb wird im Folgenden zur besseren Lesbarkeit lediglich



(a) Probe und Benennung der vier Kanten. Über den aufgebracht DMC ist jedes Teil identifizierbar. Die quadratische Kontur in der Mitte des Teils wird für die Temperaturkompensation (siehe Abschnitt 4.1.3) benötigt und spielt im weiteren Verlauf keine Rolle.



(b) Anordnung der Teile 0251 bis 0550 auf der zweiten Blechtafel in 25 Reihen (in x -Richtung der Maschine) und 12 Spalten (in y -Richtung der Maschine).

Abbildung 4.2 Probe und Anordnung der Proben auf einer Blechtafel.

diese Nummer zur Identifikation verwendet. Die Kanten des Quadrats werden nummeriert von 1 bis 4, beginnend auf der dem DMC gegenüberliegenden Kante und dann aufsteigend im Uhrzeigersinn. Dies ist in Abbildung 4.2(a) zusammengefasst. Auf eine Kante wird mit der Identifikationsnummer des Teils ergänzt durch die Kantenummer verwiesen. Die Nummer 0001-2 verweist beispielsweise auf die Kante 2 des Teils 0001.

4.1.2 Material und Schachtelung

Das verwendete Material war 3 mm dicker Edelstahl mit der Werkstoffnummer 1.4301. Die Bleche wurden kaltgewalzt und entsprachen den Normvorgaben nach DIN EN ISO 9445 [Deu10b], die Schwankungen der Dicke von bis zu $\pm 0,16$ mm erlaubt. Im schlechtesten Fall ergibt sich also eine Differenz von 0,32 mm zwischen der dünnsten und der dicksten Stelle des Blechs, was 10 % der Blechdicke entspricht. Weitere Merkmale des Blechs, z. B. chemische und mechanische Eigenschaften, entsprachen Vorgaben nach DIN EN 10088-2 [Deu14] und -4 [Deu10a] sowie DIN EN 10028-7 [Deu16].

Ein Blech hatte eine Länge von 300 cm und eine Breite von 150 cm und passte damit genau auf die Palette der Maschine (siehe Abschnitt 2.1.3). Die Quadrate wurden in 25 Reihen und 12 Spalten, wie in Abbildung 4.2(b) gezeigt, auf das Blech geschachtelt. Insgesamt waren vier Blechtafeln nötig. Alle entstammten derselben Charge. Die Teile 0001 bis 0250 wurden aus den Spalten 3 bis 12 der ersten Tafel geschnitten, die Teile 0251 bis 0550 aus der zweiten, die Teile 0551 bis 0850 aus der dritten und die Teile 0851 bis 1126 aus der vierten Tafel. Die Kanten 1 und 3 verliefen stets parallel zur y -Achse der Maschine, die Kanten 2 und 4 orthogonal dazu und damit parallel zur x -Achse. Mithilfe variabler Abstände zwischen den Teilen in x -Richtung wurde vermieden, dass die Kanten 1 oder 3 direkt über den Stegspitzen verliefen, da dies den Schneidprozess beeinflusst hätte. Ein Mindestabstand von 15 mm zwischen zwei Teilen und zum Blechrand wurde nie unterschritten. Dass die Kanten 2 und 4 an einzelnen Stellen auf einer Stegspitze auflagern, ließ sich nicht vermeiden.

4.1.3 Überwachung des Maschinenzustands

Während der mehrtägigen Abarbeitung des Versuchsplans wurde die Maschine mit einem hohen personellen und zeitlichen Aufwand überprüft und überwacht. Im normalen Produktionsumfeld ist das nicht möglich. Für diese Arbeit war es jedoch nötig, um erstens zu garantieren, dass sich der Maschinenzustand nicht verändert und es keinen Einfluss hat zu welchem Zeitpunkt ein Teil geschnitten wurde, um zweitens die Übertragbarkeit auf andere Maschinen sicherzustellen und um drittens zu gewährleisten, dass die eingestellten Soll- den tatsächlichen Ist-Parametern entsprechen.

Vor Beginn des Schneidexperiments wurden der Laser und die Maschine ausführlich überprüft und gewartet. Die Maschine wurde gereinigt und mit Schlacke verschmutzte oder beschädigte Stege wurden getauscht.

Sobald man den Laserstrahl ein- oder ausschaltet und sich die Optiken im Schneidkopf erwärmen oder abkühlen, verschiebt sich die Position des Strahlfokus entlang der Strahlachse wegen des thermisch induzierten Fokusschiffs [SM10]. Dieser wird durch Temperaturänderungen in den Optiken verursacht, weil der Brechungsindex von der Temperatur und von mechanischen Spannungen abhängt [SM10]. Es wurde vorab gemessen, wie sich die Fokusposition in Abhängigkeit von der Einschaltzeit

des Laserstrahls verschiebt. Dafür wurde unterhalb des Schneidkopfs ein BeamWatch-Messgerät der Firma Ophir Photonics platziert, das die Strahlkaustik über die Rayleigh-Streuung erfasst. In Abbildung 4.3(a) ist der Extremfall für einen ab dem Zeitpunkt 0 s dauerhaft eingeschalteten Strahl gezeigt. Als Bezugswert dient die Fokusposition einer kalten Maschine. In den ersten 10 s verschiebt sich der Fokus um ca. 170 μm . Nach 60 s hat er sich um ca. 310 μm gegenüber der Ausgangsposition bewegt. Dies ist bei den gewählten Schrittweiten von 0,5 mm für ADB und EM durchaus relevant und für die Modellierung der Zusammenhänge zwischen Fokusposition und Schnittkante äußerst hinderlich. Damit sichergestellt werden konnte, dass die eingestellte Soll- der Ist-Fokusposition entspricht, wurde dieser thermische Effekt kompensiert. Mit einer solchen Technik sind Laserschneidmaschinen mit Festkörperlaserquellen üblicherweise nicht ausgestattet, die LTT-Parameter sind aber so ausgelegt, dass sie trotz der veränderten Fokusposition einen stabilen Prozess und eine Kante guter Qualität gewährleisten. Wie in Abbildung 4.3(b) zu sehen ist, ist der Einfluss des thermisch veränderten Brechverhaltens auf den Fokusedurchmesser vernachlässigbar klein.

Bei den anderen Prozessparametern ADB, Vorschub und Gasdruck sind solche Veränderungen durch thermische Einflüsse nicht bekannt. Beim ADB wird mithilfe einer kapazitiven Regelung dafür gesorgt, dass der Soll-Abstand zum Blech auch bei einer nicht perfekt ebenen Blechoberfläche eingehalten wird.

Während des Schneidexperiments wurde der Maschinenzustand regelmäßig durch die manuelle Prüfung verschiedener Einzelkomponenten (Düse, Querblasdüse, Keramikkopf, Schutzglas, Strahlmittigkeit, Optiken, Fokus-Offset) überwacht. Zusätzlich wurden das Schutzglas und die Düse automatisch nach jedem 25. Teil kontrolliert.

4.2 Objektive Merkmale der Schnittkante

4.2.1 Messmethodik

Die quadratischen Proben haben eine Blechdicke von 3 mm und eine Kantenlänge von 10 cm, von der nur die mittleren 5 cm vermessen wurden. Zur Bestimmung objektiver Merkmale der Schnittfläche und des Grats wurde

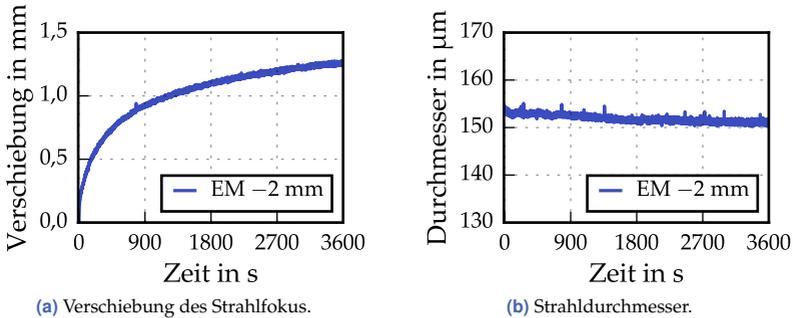


Abbildung 4.3 Verschiebung der Fokusposition und Veränderung des Strahldurchmessers ab dem Einschalten des Laserstrahls zum Zeitpunkt 0 s.

jede Schnittkante aus zwei unterschiedlichen Perspektiven erfasst. Aus jeder Perspektive wurde ein RGB-Bild und eine pixelgleiche 3D-Punktwolke aufgenommen. Das dafür verwendete Messsystem war das VR-3200 3D-Profilometer der Keyence Corporation. Dessen technische Daten sind in Anhang A.1 zu finden. Die Vermessung der vier Kanten eines Teils aus einer Perspektive dauerte ca. 15 min. Die Vermessung aller 899 Teile aus beiden Perspektiven dauerte ungefähr 450 Stunden.

4.2.1.1 Messprinzip

Das VR-3200 verwendet für die Aufnahme von RGB-Bildern rote, grüne und blaue LEDs in Kombination mit einem monochromatischen CMOS-Chip. 3D-Oberflächenprofile werden mittels Lichtschnitt erfasst. Dieses Verfahren macht sich das Triangulationsprinzip zu nutze. Lichtlinien werden auf das zu vermessende Objekt gestrahlt und das an der Oberfläche des Objekts gestreute Licht wird von einer Kamera beobachtet. Die Höhendifferenz h des Objektes entspricht dem Quotienten

$$h = \frac{\delta}{\tan(\theta)} \quad (4.1)$$

aus der lateralen Verschiebung δ der Lichtlinie (vertikal zur optischen Achse der Kamera) und dem Tangens des Triangulationswinkels θ bezüglich der optischen Achse der Kamera [BPF16]. Im VR-3200 wird für die Erzeugung der Lichtlinien eine weiße LED verwendet. Die Linien werden,

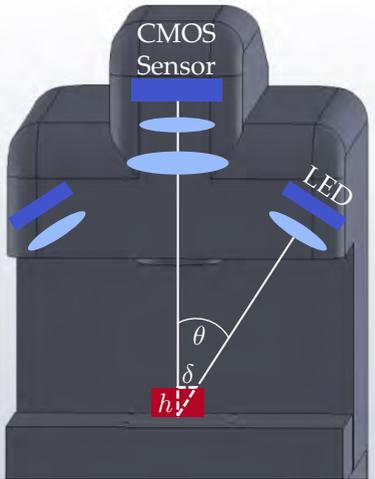
wie in Abbildung 4.4(a) zu sehen, von zwei Seiten jeweils unter dem Triangulationswinkel auf das Objekt gestrahlt. Die Erfassung der lateralen Verschiebung erfolgt mit einem doppelt telezentrischen Objektiv und dem bereits erwähnten CMOS-Chip.

4.2.1.2 Probenpositionierung

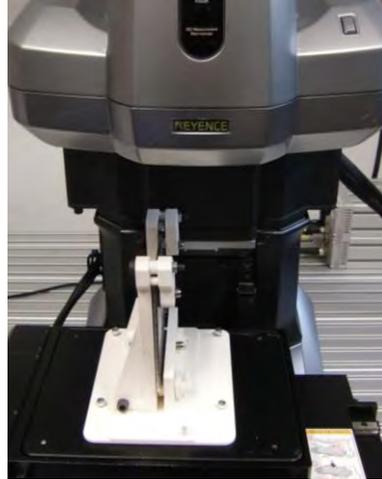
Die Perspektive des Messgeräts auf die Schnittfläche ist in den Abbildungen 4.4(b) und 4.4(c) gezeigt. Die Proben wurden so in die eigens entwickelte Halterung eingelegt, dass sich die Schnittfläche parallel zur Ebene des CMOS-Sensors befand. Die Lichtlinien verliefen entlang der Schnittkante. Für die Vermessung des Grats wurden die Proben anders positioniert, da aufgrund der unbekannt exakten Dicke des Blechs aus dem frontalen Blick auf die Schnittfläche nicht auf die Höhe des Grats geschlossen werden kann. Die Unterseite des Blechs wurde, wie in Abbildung 4.4(d) zu sehen, parallel zur Kamera ausgerichtet und diente als Referenzfläche. Die Lichtlinien verliefen quer zum Grat. Es wäre vorteilhaft gewesen, die Blechunterseite bei der Gratsmessung nicht parallel, sondern schräg zur Kamera auszurichten, weil dann nicht nur die oberste Spitze des Grats, sondern die gesamte Flanke sichtbar gewesen wäre. Aufgrund der geringen Tiefenschärfe von 1 mm ($\pm 0,5$ mm) war dies jedoch nicht möglich.

4.2.1.3 Messfehler

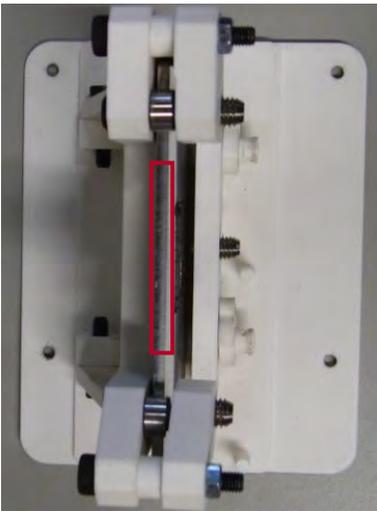
Bei der Erfassung der 3D-Punktwolken mancher Schnittflächen kommt es zu Fehlern. Ein typisches Beispiel für das beschriebene Problem ist in Abbildung 4.5 zu sehen. Dort ist ein zweimal vermessener Ausschnitt der Kante 0012-4 gezeigt. Während die RGB-Bilder beider Messungen sehr ähnlich aussehen (siehe Abbildung 4.5(a) und 4.5(c)), unterscheiden sich die 3D-Punktwolken, insbesondere im Bereich der Blechoberkante deutlich. Bei der Messung I treten tiefe Löcher auf, bei der Messung II sind diese verschwunden. Die stichprobenartige Untersuchung einiger Kanten mit einem Mikroskop zeigte, dass es sich bei den tiefen Löchern stets um Messfehler handelt. Die in Abbildung 4.5 gezeigte Kante wurde bei beiden Messdurchläufen in dieselbe Probenhalterung eingespannt und mit denselben Einstellungen vermessen. Die minimal andere Positionierung der



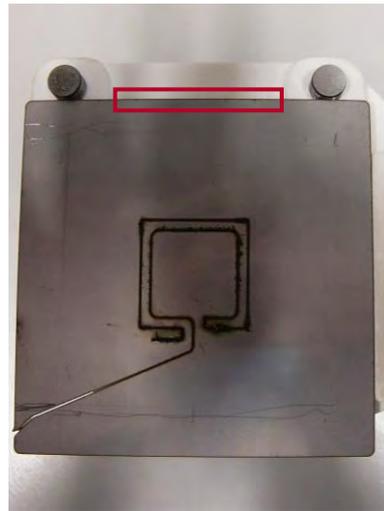
(a) Messprinzip Lichtschnitt.



(b) Probenpositionierung Schnittfläche.



(c) Perspektive auf die Schnittfläche.



(d) Perspektive auf den Grat.

Abbildung 4.4 Messprinzip im VR-3200 Messgerät, Probenpositionierung bei der Aufnahme der Schnittfläche und Perspektive des Messgeräts bei der Aufnahme der Schnittfläche und des Grats. Der erfasste Bereich ist in Abbildung (c) bzw. (d) jeweils rot eingrahmt.

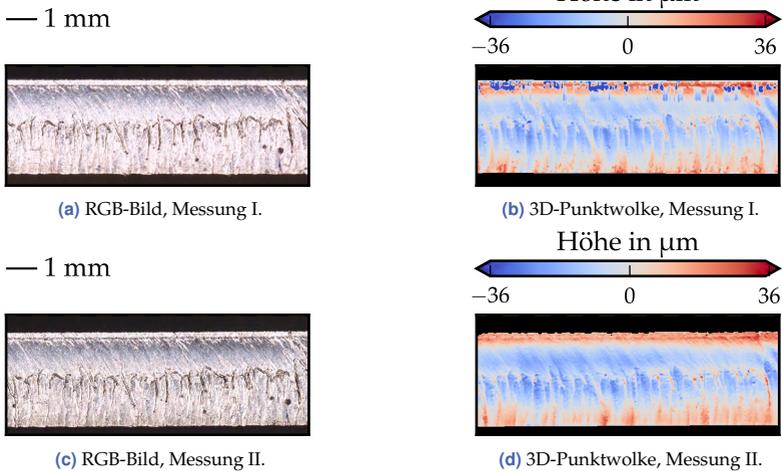


Abbildung 4.5 RGB-Bilder und 3D-Punktwolken der Kante 0012-4 für zwei verschiedene Messdurchgänge. Bei Messdurchgang I treten Fehler auf.

Probe durch das erneute Einlegen führt dazu, dass der Fehler verschwindet. In Abschnitt 6.1.2 des Ergebniskapitels werden die Gründe für diese Beobachtung diskutiert und das Problem wird detailliert analysiert.

Bei den Falschfarbendarstellungen der 3D-Punktwolken ist zu beachten, dass in der Punktwolke Werte enthalten sein können, die außerhalb von den Grenzen der Farbskala liegen. Dies ist durch die pfeilförmigen Abschlüsse der Skala angedeutet und geschieht zu Gunsten der Auflösung kleinerer Werte. Kleinere bzw. größere Werte sind einheitlich dunkelblau bzw. dunkelrot eingefärbt. In Abbildung 4.5(b) trifft dies zum Beispiel auf die Löcher im oberen Bereich zu.

Auf die weitere Verwendung der 3D-Messdaten haben diese Fehler große Auswirkungen. Hiervon ist insbesondere die Rautiefenmessung betroffen, die auf der Höhendifferenz zwischen der höchsten Spitze und der tiefsten Senke basiert (siehe Abschnitt 2.2.1.1). Die tiefen Senken verursachen hohe Rautiefenwerte. Dies verschlechtert die Datenqualität deutlich und wird im weiteren Verlauf der Arbeit immer wieder thematisiert.

4.2.2 RGB-Bild und 3D-Punktwolke

Mit einer Aufnahme kann ein $7,6 \text{ mm} \times 5,7 \text{ mm}$ großer Bereich erfasst werden. Für die Aufnahme der 5 cm langen Schnittkante werden mehrere Aufnahmen zusammengesetzt (engl. *stiching*), sodass ein $7,6 \text{ mm} \times 50,0 \text{ mm}$ großes Bild entsteht. Die Bildgröße in Pixel beträgt 521×3400 . Ein Pixel entspricht also einem Bereich von ca. $15 \mu\text{m} \times 15 \mu\text{m}$. Das RGB-Bild und die 3D-Punktwolke sind pixelgleich. Ersteres hat eine Farbtiefe von 8 bit pro Kanal. Die Pixel der 3D-Punktwolke beinhalten positive und negative Höhenwerte in μm . Pixel mit positiven Werten liegen oberhalb des Höhen-Mittelwerts, Pixel mit negativen Höhenwerten befinden sich unterhalb. Die Genauigkeit der Höhenmessung beträgt $\pm 3 \mu\text{m}$ und die laterale Genauigkeit liegt bei $\pm 2 \mu\text{m}$.

4.2.2.1 Vorverarbeitung der RGB-Bilder und 3D-Punktwolken für CNNs

Die Aufnahmen des Grats wurden nicht als Eingangsdaten für CNNs genutzt. Die RGB-Bilder und 3D-Punktwolken der Schnittfläche wurden wie folgt vorverarbeitet.

Die Bilder wurden zunächst zugeschnitten, um den Hintergrund zu entfernen. Hierfür wurde der Canny-Algorithmus [Can86] verwendet, mit dem die Ober- und Unterkante der Schnittkante im RGB-Bild identifiziert werden konnten. Anschließend wurde der Hintergrund entfernt. Die Oberkante wurde dann genutzt, um die Schnittkante möglichst horizontal im Bild auszurichten. Für alle in dieser Arbeit gezeigten Schnittkanten gilt, dass die Oberseite des Blechs auch in der Abbildung oben ist. Die Schnittrichtung war bei allen Kanten von rechts nach links.

Das Zuschneiden und horizontale Ausrichten konnten anschließend auf die 3D-Punktwolke übertragen werden, da beide Aufnahmen pixelgleich sind. Die Höhenwerte wurden zusätzlich mithilfe einer Referenzebene korrigiert, um den Einfluss der Positionierung der Probe zu eliminieren. Dazu wurde in Schnittrichtung (lange Bildseite) eine Ebene mithilfe einer linearen Regression in die gemessene Oberfläche gefittet und anschließend abgezogen. Quer zur Schnittrichtung (kurze Bildseite) wurde keine Korrektur ausgeführt, um keinen Einfluss auf die Rechtwinkligkeitstoleranz zu nehmen. Danach wurde die 3D-Punktwolke in ein Graustufenbild mit einer Farbtiefe von 8 bit überführt. Um die Auflösung im relevanten



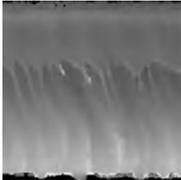
Abbildung 4.6 Vorverarbeitung der Bilder zur Verwendung als Eingangsdaten für neuronale Netze. Zuerst wird die Schnittkante des hier gezeigten Originalbildes freigestellt und horizontal ausgerichtet. Dann wird sie in 15 quadratische Bilder zerschnitten (von denen hier drei gezeigt sind). Die quadratischen Bilder werden auf die einheitliche Größe von 224×224 Pixeln skaliert.

Bereich zu maximieren, wurden Werte, die mehr als das Dreifache des Interquartilsabstands unterhalb des 0,25-Quantils bzw. oberhalb des 0,75-Quantils lagen, durch den Median ersetzt. Der kleinste Höhenwert wurde auf 1 gesetzt, der größte auf 255. Fehlende Werte erhielten den Wert 0. Dieses Graustufenbild wird im Folgenden auch als Höhenbild bezeichnet.

Die Bilder haben – unabhängig vom Typ – dasselbe längliche Format mit den ungefähren Abmessungen 215×3400 Pixel (abhängig von der Grathöhe) und bestehen aus über 700000 Pixeln. Für das Training neuronaler Netze ist dieses Format äußerst unüblich, außerdem steigt der Rechenaufwand mit der Neuronenanzahl. Gleichzeitig beinhaltet ein solches Bild viel redundante Information, weil sich das Aussehen in Schnittrichtung entlang der gesamten Kante kaum verändert. Deshalb wurde jedes Bild in 15 quadratische Teile geschnitten, die anschließend auf die Größe 224×224 Pixel skaliert wurden. Dies ist in Abbildung 4.6 für ein RGB-Bild zu sehen. Für jede Kante liegen 15 und für jedes Teil 60 Bilder eines Typs vor.

RGB und Höhenbild wurden zu einem dritten Typ, einem RGBD-Bild (von engl. *depth*, Tiefe) mit vier Kanälen, zusammengesetzt. Dieses enthält sowohl die Farb- als auch die Höheninformation. Ein Beispiel für ein RGB-, ein zugehöriges Höhen- und RGBD-Bild ist in Tabelle 4.2 abgebildet.

Tabelle 4.2 Beispiel für ein RGB-, ein Höhen- und ein RGBD-Bild. Das RGBD-Bild hat vier Kanäle und ist hier als RGBA-Bild dargestellt. Die Höhe bestimmt in dieser Darstellung die Transparenz der RGB-Pixel.

Bildtyp	RGB	Höhe	RGBD
			
Kanäle	3	1	4

4.2.3 Objektive Merkmale der Schnittfläche

Aus den gemessenen 3D-Punktwolken von Schnittfläche und Grat wurden die Rautiefe und die Rechtwinkligkeitstoleranz sowie die Grathöhe und Gratbreite abgeleitet. Ausgangsdaten dafür waren die 3D-Rohdaten (und nicht die im vorhergehenden Abschnitt beschriebenen Graustufenbilder).

4.2.3.1 Rautiefe

Zur Bestimmung der Rautiefe mit der VR-3200 Software wurden die Rohdaten zunächst mit einer Referenzebene korrigiert, um den Einfluss der Probenpositionierung zu eliminieren. Anschließend wurden Messlinien in Schnittrichtung in verschiedenen Tiefen des Blechs definiert. Die oberste Messlinie hat einen Abstand von 0,3 mm von der Oberkante, die weiteren Messlinien befinden sich in 0,3 mm-Schritten darunter. Die unterste Messlinie hat einen Abstand von 2,7 mm von der Ober- und ca. 0,3 mm von der Unterkante. Die Messlinien werden entsprechend ihrer Position mit 03, 06, 09 usw. bezeichnet. Im Folgenden muss zwischen der Messung nach Norm und der in dieser Arbeit alternativ verwendeten Methode unterschieden werden.

Messung nach DIN EN ISO 4288 [Deu98b] Wie in Abschnitt 2.2.1.1 erklärt und in der Tabelle 2.2 zusammengefasst, muss bei der Messung nach Norm für die zu erwartenden Rautiefenwerte eine Messstrecke von

4,0; 12,5 oder 40,0 mm Länge mit einer Grenzwellenlänge von 0,8; 2,5 oder 8,0 mm ausgewertet werden. Da ein 50 mm langer Bereich der Schnittkante zur Verfügung steht, kann in jeder Blechtiefe nur einmal gemessen werden. Je nach Wertebereich von $Rz5$ ist nur eine der drei Messungen gültig. Die anderen zwei werden verworfen. So erhält man zu jeder Blechtiefe genau einen $Rz5$ -Wert. Da nach DIN EN ISO 9013 [Deu17] nur der maximale Rautiefenwert interessant ist, ergibt sich ein $Rz5$ Wert pro Schnittkante.

Diese Messmethode hat mehrere Nachteile: Es kommt zu inkonsistenten Werten, weil der Wechsel der Grenzwellenlänge (und damit der Eigenschaften des Filters, das die kurz- von den langwelligen Anteilen trennt) zwischen verschiedenen Rautiefenbereichen zu Sprüngen führt. Nachteilig ist außerdem, dass nur eine Messung für jede Blechtiefe möglich ist. Es liegen also keine Vergleichswerte in derselben Blechtiefe vor, was eine Ausreißerdetektion erschwert. Da nur der maximale Wert angegeben wird, fallen die Ausreißer zusätzlich ins Gewicht. Die Angabe eines $Rz5$ -Wertes spiegelt das tatsächliche Aussehen der Schnittkante, die oft mehrere unterschiedlich raue Schichten aufweist, nicht wider. Diese Messmethode wird im weiteren Verlauf der Arbeit daher nicht mehr angewendet.

Messung in fünf Bereichen Aufgrund der beschriebenen Nachteile der normkonformen Rautiefenmessung wurde eine an die Norm angelehnte, aber etwas robustere Methode entwickelt. Streng genommen ist die Messung der Rautiefe mit einem optischen Verfahren sowieso nicht normkonform, da ein Tastverfahren empfohlen wird. Aus Zeitgründen kam dies für die Vermessung aber nicht infrage.

Der aufgenommene Bereich der Schnittkante wurde, wie in Abbildung 4.7 zu sehen ist, in fünf gleich große Abschnitte geteilt. In jedem Bereich wurden neun Messlinien in verschiedenen Blechtiefen wie oben beschrieben platziert. Es wurde eine feste Grenzwellenlänge von 2,5 mm gewählt. Anstatt des $Rz5$ -Wertes muss jeweils der $Rz3$ -Wert berechnet werden, weil aufgrund der Bereichsgröße von 10 mm nur über drei statt fünf Einzelmessstrecken gemittelt werden kann. Auf diese Weise erhält man für jede Schnittfläche statt eines $Rz5$ -Wertes nach Norm 45 $Rz3$ -Werte, fünf für jede der neun Blechtiefen. Die einzelnen Rautiefenwerte werden mit

$$Rz_{M_L}^{(M_B)} \text{ mit } M_L \in \{03, 06, \dots, 27\} \text{ und } M_B \in \{1, 2, 3, 4, 5\} \quad (4.2)$$

bezeichnet, wobei M_L die Messlinie und M_B den Messbereich benennt. Der große Vorteil dieser Methode ist, dass die größere Anzahl an Messstellen die Detektion und Korrektur von Messfehlern ermöglicht. Die Datenqualität kann z. B. quantifiziert werden, indem die absoluten

$$\delta_{M_L}^{(M_B)} = Rz_{M_L}^{(M_B)} - \min_{M_B} \left(Rz_{M_L}^{(M_B)} \right) \quad (4.3)$$

oder die relativen Differenzen

$$\delta_{M_L}^{(M_B)} = \frac{Rz_{M_L}^{(M_B)} - \min_{M_B} \left(Rz_{M_L}^{(M_B)} \right)}{\min_{M_B} \left(Rz_{M_L}^{(M_B)} \right)} \quad (4.4)$$

der fünf $Rz3$ -Werte einer Messlinie M_L betrachtet werden. Da die fünf Bereiche einer Kante mit denselben Prozessparametern und (fast) unter denselben Umgebungsbedingungen geschnitten wurden, sollte sich die Rautiefe kaum unterscheiden. Im Allgemeinen ist diese Differenzbildung sicherlich keine geeignete Methode, um Ausreißer zu finden, da das Minimum selbst empfindlich von Ausreißern abhängen kann. Die in Abschnitt 4.2.1.3 und 6.1.2 beschriebenen Messprobleme führen jedoch stets zu einer Erhöhung des Rautiefenwerts. Außerdem können basierend auf fünf Werten keine verlässlichen statistischen Kennzahlen, wie beispielsweise die Standardabweichung, ausgerechnet werden. Deshalb wird das Minimum der fünf Messwerte als verlässlichster Wert eingestuft.

4.2.3.2 Rechtwinkligkeitstoleranz

Analog zur Messung der Rautiefe wird auch die Rechtwinkligkeitstoleranz u in fünf Bereichen der Schnittfläche ermittelt. In jedem Bereich wird dafür eine Messlinie verwendet, die orthogonal zu den Rautiefenmesslinien verläuft. Wie in Abschnitt 2.2.1 beschrieben, wird nur der Bereich ab 0,3 bis 2,7 mm unterhalb der Blechoberkante vermessen, damit Verformungen der Ober- und Unterkante nicht eingehen. Daraus resultieren fünf Werte $u^{(M_B)}$ pro Kante.

4.2.4 Objektive Merkmale des Grats

Für die Ermittlung der Gratmerkmale wird eine aus einer anderen Perspektive (siehe Abbildung 4.4(d)) aufgenommene 3D-Punktwolke heran-

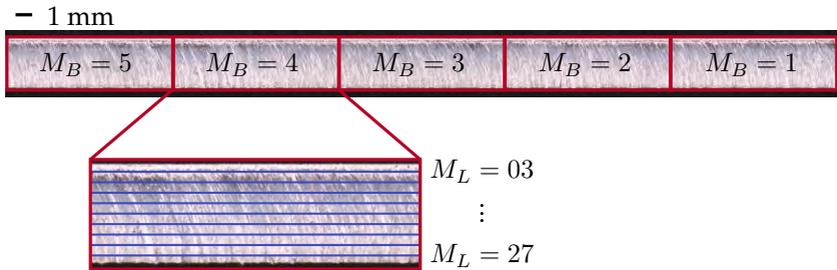


Abbildung 4.7 Aufteilung der Schnittfläche in fünf Bereiche M_B mit Nummerierung in Schnittrichtung und Definition von neun Messlinien M_L in jedem Bereich. Diese werden entsprechend ihres Abstands von der Blechoberkante mit 03 (0,3 mm) bis 27 (2,7 mm) bezeichnet.

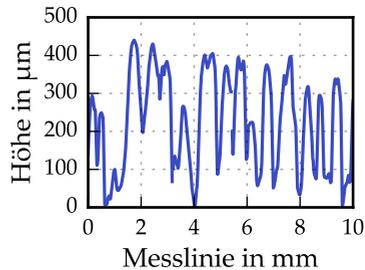
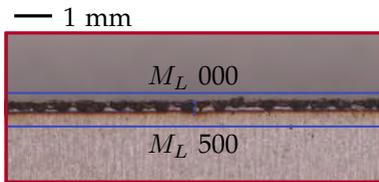
gezogen. Da die Gratperlen meist dunkel verfärbt sind, ist nicht davon auszugehen, dass es zu den in Abschnitt 4.2.1.3 beschriebenen Messfehlern kommt.

Analog zur Rautiefenmessung wird der aufgenommene Bereich der Schnittkante in fünf Teilbereiche aufgeteilt. Die Norm DIN EN ISO 9013 [Deu17] macht keinerlei Vorgaben zur Vermessung des Grats. In dieser Arbeit wurden zwei Eigenschaften identifiziert, die besonders interessant sein könnten, die Höhe und die Breite des Grats.

Diese Merkmale können mit dem VR-3200 nicht direkt erfasst werden, sondern wurden mithilfe der in Abbildung 4.8 gezeigten Profilmessungen konstruiert. Es wurden 26 Profilmesslinien im Abstand von $20\ \mu\text{m}$ parallel zur Schnittrichtung auf den Grat gelegt. In Abbildung 4.8(a) sind die beiden äußersten Messlinien eingetragen, die anderen 24 liegen dazwischen. Analog zur Rautiefe werden sie entsprechend ihrer Position benannt mit 000, 002, 004 usw. Jede Messlinie erfasst die Höhe des Grats bezüglich der Blechunterseite. Ein Beispiel ist in Abbildung 4.8(b) gezeigt.

Zur Ermittlung der Grathöhe wird aus jeder Messlinie der größte Höhenwert (im Beispiel $440\ \mu\text{m}$) extrahiert. Der maximale Höhenwert der 26 Profilmesslinien wird als Grathöhe definiert.

Für die Gratbreite werden alle Messlinien betrachtet, die einen Höhenwert größer als $10\ \mu\text{m}$ aufweisen. Der Abstand zwischen der ersten und der letzten Messlinie, für die dies ohne Unterbrechungen zutrifft, wird Gratbreite genannt. Man erhält so einen Wert für die Grathöhe und einen



(a) Messlinien in einem der fünf Messbereiche.

(b) Beispiel für eine Profilmessung.

Abbildung 4.8 Vermessung des Grats mit Profilmesslinien. Die Messlinien werden entsprechend ihres Abstands von der ersten Linie mit 000 (0 mm) bis 500 (5 mm) benannt.

für die Gratbreite in jedem der fünf Messbereiche. Analog zur Rautiefe können diese genutzt werden, um eventuelle Mess- oder Prozessfehler zu detektieren.

4.3 Ausreißerkorrektur

Für jede Schnittkante liegen 45 Rautiefenmesswerte vor. Diese sind wie in Abschnitt 4.2.1.3 beschrieben teilweise fehlerbehaftet. Da sich dies negativ auf die Güte der datenbasierten Modelle auswirken könnte, wurden drei Methoden zur Verbesserung der Datenqualität evaluiert.

Kombination zweier Messdurchgänge Eine Möglichkeit, die Qualität der Messdaten zu verbessern, ist es, mehrere Messungen durchzuführen und die Messergebnisse zu kombinieren. Die Wiederholung der Messung ist sehr zeitaufwendig und wurde deshalb nur einmal durchgeführt. Da nur Messfehler bekannt sind, die zu einer Erhöhung des Rz -Wertes führen, wurden die Werte der beiden Messdurchgänge komponentenweise (selbe Kante, selbe Messlinie, selber Bereich) verglichen und es wurde der jeweils kleinere Wert übernommen.

Jede Schnittkante mehrmals in die Probenhalterung einzulegen und zu vermessen, ist sicherlich der zuverlässigste Weg, um die Datenqualität zu verbessern. Dies ist jedoch sehr zeitaufwendig. Darum wurden zwei Ansätze evaluiert, mit denen die Datenqualität basierend auf einem

Messdurchgang verbessert werden kann. Dafür musste die Annahme getroffen werden, dass die Rautiefenwerte einer Messlinie im Idealfall nicht voneinander abweichen. Dies soll sowohl für die fünf Werte innerhalb einer Kante als auch näherungsweise für die 20 Rautiefenwerte eines Teils gelten, da diese mit denselben Prozessparametern erzeugt wurden und sich nur hinsichtlich der Walz- und Schnittrichtung unterscheiden. Für die Korrektur wurden bewusst eine sehr einfache Methode mit einem statistischen Schwellenwert und ein sehr fortgeschrittener Ansatz mit einem Autoencoder gewählt.

Std-Schwelle und Mittelwertkorrektur Der Mittelwert und die Standardabweichung (Std) der 20 Rautiefenwerte jeder Messlinie werden berechnet. Weicht ein Rz -Wert um mehr als einen Schwellenwert vom Mittelwert ab, wird dieser als Ausreißer betrachtet. Der Schwellenwert entspricht dem 1,7-fachen der Standardabweichung. Die Begründung dafür ist in Abschnitt 6.1.2.2 zu finden. Ausreißer werden anschließend durch den (neu berechneten) Mittelwert ersetzt.

Alternativ zu den oben beschriebenen Methoden wurde ein Autoencoder zur Detektion und Korrektur der Messdaten verwendet. Dieser wird im Folgenden näher thematisiert.

4.3.1 Korrektur mit einem Autoencoder

Autoencoder sind spezielle künstliche neuronale Netze. Anders als bei den meisten Anwendungen von KNNs wird $f(x) = x$ anstatt $f(x) = y$ berechnet [Agg17]. Beim Training wird der Abstand zwischen dem Original am Eingang und der sogenannten Rekonstruktion am Ausgang minimiert. Die generelle Struktur eines Autoencoders umfasst einen Encoder und einen Decoder, zwischen denen sich der sogenannte Flaschenhals (engl. *bottleneck*) befindet. Dieser umfasst meistens deutlich weniger Neuronen als die Ein- und Ausgangsschicht. Der Encoder bildet die am Eingang übergebene Information auf eine komprimierte Repräsentation im Flaschenhals ab. Der Decoder verwendet die komprimierte Information, um eine Rekonstruktion des Eingangs zu erstellen. Oft (aber nicht notwendigerweise) handelt es sich bei Encoder und Decoder um symmetrisch aufgebaute neuronale Netze.

Es gibt Anwendungen, bei denen man sich für die Neuronen des Flaschenhalses interessiert, die die komprimierte Information beinhalten. Autoencoder können also zur Dimensionsreduktion verwendet werden. Sie generieren häufig bessere Repräsentationen als eine Hauptkomponentenanalyse oder eine Matrix-Faktorisierung [Hin06]. Der Hauptgrund dafür ist, dass sich mit nicht linearen neuronalen Netzen beliebige Datenverteilungen besser modellieren lassen. Es ist davon auszugehen, dass sie bei der Detektion von Ausreißern ebenfalls besser abschneiden. Bei der Ausreißerdetektion werden nicht die Neuronen des Flaschenhalses verwendet, sondern die Rekonstruktion am Ausgang wird komponentenweise mit dem Eingang verglichen. Da die komprimierte Information im Flaschenhals des Autoencoders robust gegenüber Rauschen und Ausreißern ist, werden diese am Ausgang nicht rekonstruiert. Deshalb deuten große Unterschiede zwischen Eingang und Rekonstruktion auf Ausreißer in den Eingangsdaten hin. [Agg17]

Aggarwal [Agg17] schreibt, dass die Verwendung neuronaler Netze zur Ausreißerdetektion in der Praxis herausfordernd ist, weil diese im Training empfindlich auf Rauschen und Ausreißer reagieren. Da Ausreißer während des Trainingsprozesses als normale Datenpunkte behandelt werden, beeinflussen sie die zu minimierende Verlustfunktion. Er rät deshalb, mehrere Modelle jeweils auf Teilmengen der zur Verfügung stehenden Daten zu trainieren und die Ausreißerdetektion anschließend nur für unbekannte Daten durchzuführen. Eine andere Möglichkeit, die Robustheit zu erhöhen, sei es, Dropout zu verwenden.

Vincent et al. [Vin+10] schlagen eine andere Herangehensweise vor, die in dieser Arbeit eingesetzt wird: Der Autoencoder wird nicht mithilfe der Originaldaten trainiert, sondern mit einer zusätzlich stark verfälschten Version \tilde{x} . Dazu werden in einem ersten Schritt in jedes Trainingsbeispiel viele weitere Ausreißer eingefügt und anschließend wird Dropout angewendet. Diese verfälschte Information wird während des Trainings am Eingang des Autoencoders übergeben. Die unverfälschten Originaldaten werden als Label genutzt. Das Ziel des Trainings ist es, die originalen Daten möglichst gut aus den verfälschten zu rekonstruieren: $f(\tilde{x}) = x$. So wird eine gewisse Unabhängigkeit von der Datenqualität der Originaldaten erreicht und das Modell lernt, Rauschen und Ausreißer zu entfernen bzw. zu korrigieren, anstatt nur eine identische Abbildung durchzuführen.

Das so trainierte Modell kann anschließend zur Ausreißerkorrektur eingesetzt werden. In diesem Fall werden die Originaldaten ohne zusätzliche Verfälschung übergeben.

4.3.1.1 Verfälschung der Rautiefendaten für das Modelltraining

Die originalen Rautiefenmesswerte werden als Tensor mit den Dimensionen $5 \times 9 \times 4$ an den Autoencoder übergeben. Durch diese Struktur bleibt die räumliche Information der Daten (fünf Bereiche, neun Messlinien, vier Kanten) erhalten. Um den Autoencoder zu trainieren, wird die Eingangsdatenqualität massiv verschlechtert. Zuerst werden zusätzliche (zufällige) Ausreißer in die Originaldaten eingefügt. Um negative oder zu große Werte zu vermeiden, werden Ausreißer, die außerhalb des Rz -Bereichs der Originaldaten (4 bis $534 \mu\text{m}$) liegen, auf die Grenzen dieses Bereichs gesetzt. Die Wahrscheinlichkeit für das Einfügen solcher zusätzlicher Ausreißer beträgt 20 %. Im Schnitt werden also 36 der 180 Eingangsneuronen auf diese Weise manipuliert. In einem zweiten Schritt entfernt eine Dropout-Schicht 30 % aller Eingangswerte. Anschließend werden diese stark verrauschten Daten an den Autoencoder übergeben. Die zu minimierende Verlustfunktion im Training ist der mittlere absolute Fehler (siehe Formel (3.7)) zwischen den originalen Daten als Label und der Repräsentation.

4.3.1.2 Konfiguration des Autoencoders

Die Architektur des verwendeten Autoencoders ist in Abbildung 4.9 veranschaulicht. Encoder und Decoder bestehen jeweils aus drei Highway-Blöcken (siehe Abschnitt 3.2.4) und einer Faltungsschicht zur Dimensionsreduktion im Flaschenhals. Innerhalb eines Highway-Blocks wenden die Faltungsschichten H und T jeweils vier Filter der Größe 3×3 auf ihre Vorgängerschicht an. Jede dieser Schichten hat 148 trainierbare Parameter. H benutzt ReLU als Aktivierungsfunktion, bei T kommt die logistische Funktion zum Einsatz (siehe Abschnitt 3.2.1.2). Im Flaschenhals wird ebenfalls eine Faltung mit einem Filter derselben Größe verwendet. Dort wird die Anzahl der Neuronen auf ein Viertel reduziert. Encoder und Decoder sind symmetrisch aufgebaut. Mithilfe von Zero Padding (siehe Abschnitt 3.2.3) wird eine Verringerung der räumlichen Dimension bei der

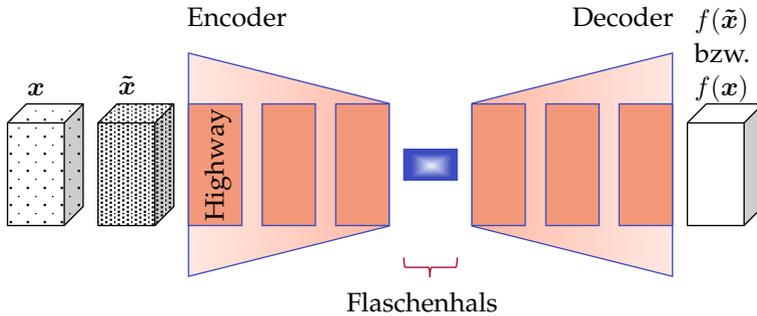


Abbildung 4.9 Architektur des Autoencoders zur Ausreißerkorrektur in den Rautiefenmessdaten. Encoder und Decoder bestehen jeweils aus drei Highway-Blöcken. Für das Training wird am Eingang eine verfälschte Version \tilde{x} der Originaldaten x übergeben. Entsprechend muss die Repräsentation in diesem Fall mit $f(\tilde{x})$ bezeichnet werden. Für die Bereinigung der Messfehler wird der trainierte Autoencoder eingesetzt: x wird übergeben und die berechnete Repräsentation $f(x)$ wird weiterverwendet.

Faltung verhindert. Im Anschluss an jeden Highway-Block und die Faltungsschichten im Flaschenhals sorgen eine Dropout- und eine BN-Schicht (siehe Abschnitt 3.2.3) für einen effektiveren Trainingsprozess. Insgesamt hat der Autoencoder 2117 Modellparameter, von denen 2059 während des Trainingsprozesses (mit dem Adam-Optimierer [BK15]) optimiert werden.

4.3.1.3 Ausreißerkorrektur

Nach dem Training kann der Autoencoder für die Korrektur der Messdaten verwendet werden. Dazu werden die originalen Daten an das Modell übergeben, welches daraus eine Rekonstruktion berechnet. Große Unterschiede zwischen Originaldaten und Rekonstruktion deuten auf Ausreißer hin. Damit nicht alle Werte geändert werden, was zur Folge hätte, dass die gesamten im weiteren Verlauf der Arbeit verwendeten Rautiefendaten durch ein neuronales Netz verändert wären, wird der ursprüngliche Wert $x_{\text{orig.}}$ nur durch den neu berechneten $x_{\text{repr.}}$ ersetzt, falls gilt:

$$\frac{|x_{\text{orig.}} - x_{\text{repr.}}|}{x_{\text{orig.}}} > 0,35. \quad (4.5)$$

Falls diese Bedingung nicht erfüllt ist, wird der ursprüngliche Wert beibehalten. Die Begründung für die Wahl einer Schwelle von 0,35 ist in Abschnitt 6.1.2.2 zu finden.

4.4 Konzept der Expertenbefragung

In der Praxis wird die Schnittkantenqualität üblicherweise nicht anhand objektiver Messwerte beurteilt, sondern subjektiv durch den Menschen. In dieser Arbeit wurde eine Expertenbefragung durchgeführt, um einerseits zu untersuchen, inwieweit eine einzelne Person in der Lage ist, ein repräsentatives Qualitätsurteil zu fällen und um andererseits zu verstehen, welche Eigenschaften der Schnittkante für den subjektiven Eindruck besonders wichtig sind.

Bei den befragten fünf Personen handelte es sich um TRUMPF-Versuchsmitarbeiter mit langjähriger Erfahrung im Laserschneiden. Laien kommen für eine derartige Aufgabe nicht infrage. Aus Zeitgründen konnten nicht alle geschnittenen Kanten beurteilt werden. Stattdessen wurden 100 Schnittkanten ausgewählt und verteilt über drei Wochen bewertet.

In Ermangelung eines anderen Kriteriums wurde die gemessene Rautiefe (siehe Abschnitt 4.2.3.1) der Kanten verwendet, um aus den 3596 Schnittkanten 100 repräsentative auszuwählen. Dafür wurde die Bandbreite der Rautiefe so in zehn Intervalle aufgeteilt, dass jeder Bereich gerade 10 % der Schnittkanten enthielt. Anschließend wurden aus jedem Bereich zufällig zehn Kanten ausgewählt.

Jede Woche wurden 40 Schnittkanten beurteilt, dabei waren 30 Kanten neu, zehn Kanten wurden in jeder der drei Wochen ohne das Wissen der Befragten wiederholt. Es wurde dabei nur die gerade Schnittkante bewertet, die Ecken waren abgedeckt. Die Experten durften die Teile in die Hand nehmen und die Kanten anschauen und anfassen, wie sie es gewohnt sind. Messmittel standen keine zur Verfügung. Jeder Experte musste jede Kante mit zwei Schulnoten (von 1 „sehr gut“ bis 6 „ungenügend“) bewerten. Eine Note sollte für die Qualität der Schnittfläche und eine für die Qualität des Grats vergeben werden.

In Woche 1 sollten die Experten zusätzlich zu der Schulnote den Hauptgrund für ihre Bewertung der Schnittfläche als Freitext angeben. Die Antworten wurden in fünf Kategorien zusammengefasst: Rauheit, Rechtwink-

Tabelle 4.3 Übersicht über die Datenbasis.

Parameterkombinationen	1126
Unterschiedliche Parameterkombinationen	1050
Vollfaktoriell variierte Prozessparameter	4
Fehlschnitte	227
Verbleibende Teile	899
Davon mit verschiedenen Parameterkombinationen	834
Kanten pro Teil	4

lichkeit, Verfärbung, Homogenität und Verrundung der Oberkante. In den darauffolgenden Wochen sollten diese Kategorien zusätzlich zur Benotung bewertet werden. Dabei gab es nur die Auswahlmöglichkeiten gut, mittelmäßig und schlecht.

4.5 Übersicht über die Datenbasis

Als kurze Zusammenfassung des Kapitels ist in Tabelle 4.3 eine Übersicht über die Datenbasis gezeigt. Der Versuchsplan bestand aus 1126 Parameterkombinationen, von denen 1050 verschieden waren. 227 Parameterkombinationen führten zu Fehlschnitten, die Teile konnten nicht aus dem Restgitter entnommen werden. 899 Teile konnten entnommen und weiter verwendet werden. Jedes Teil hat vier Kanten. Es stehen also insgesamt 3596 Kanten zur Verfügung. Die Datenbasis für das Training der CNNs enthält über 50000 Bilder, weil es zu jedem Teil 60 quadratische Bilder gibt.

Jede Kante kann anhand einer Identifikationsnummer identifiziert werden. Diese setzt sich zusammen aus der Identifikationsnummer des Teils, einer Zahl zwischen 0001 und 1126, und der angehängten Nummer der Kante, -1, -2, -3 oder -4. In Tabelle 4.4 ist gezeigt, welche Daten für jede Kante verfügbar sind.

Tabelle 4.4 Verfügbare Daten zu jeder der 3596 Kanten. (*Die subjektiven Merkmale sind nur für 100 Kanten vorhanden.)

Prozessparameter	Abstand Düse-Blech Einstellmaß Gasdruck Vorschub
Zusätzlich	Fokusposition
Prozess-, Maschinendaten	Blechnummer Datum und Uhrzeit Position auf dem Blech Maschinenzustand
15 quadratische Bilder	RGB-Bild Graustufen-Höhenbild RGBD-Bild (Farbe und Höhe)
Objektive Merkmale	45 Rz -Werte 5 u -Werte 5 Werte für die Grathöhe 5 Werte für die Gratbreite
Subjektive Merkmale*	Schulnote für die Schnittfläche Schulnote für den Grat Freitext/Bewertung von fünf Kategorien

5 Modellierung und Visualisierung

Dieses Kapitel baut auf den in Kapitel 3 eingeführten Grundlagen des maschinellen Lernens auf. Sofern nicht anders angegeben sind alle Fachbegriffe in Abschnitt 3.2.1 zu finden. Zu Beginn werden die CNN-Architekturen und -Konfigurationen, die zur Modellierung der Zusammenhänge zwischen Schnittkante und Prozessparametern sowie zur Schätzung der Rautiefe verwendet wurden, erläutert. Daran schließt die Herleitung und Erklärung der Layer-wise Relevance Propagation an. Mit dieser Visualisierungstechnik können die von einem CNN erlernten Muster und Zusammenhänge dargestellt werden. Danach wird kurz auf die Ansätze eingegangen, die zur Modellierung der Zusammenhänge zwischen Prozessparametern und Fehlschnitttrisiko, Rautiefe oder Grathöhe genutzt wurden. Am Ende des Kapitels werden alle Modellansätze in Tabelle 5.3 zusammengefasst.

5.1 Bildbasierte Modellierung mit neuronalen Netzen

5.1.1 Regression der Prozessparameter

Für die automatisierte Überprüfung des Zustands einer Laserschneidmaschine benötigt man ein Modell, welches die zugrunde liegenden Prozessparameter ausgehend von (einem Bild) der Schnittkante schätzt. Mehr zu dieser Idee befindet sich in Abschnitt 6.3. Zunächst waren bekannte Netzarchitekturen (z. B. VGG-Netz [SZ15], AlexNet [KSH17], ResNet [He+16]) auf ihre Anwendbarkeit für diese Aufgabe getestet worden. Aufgrund der großen Anzahl zu trainierender Modellparameter wurden vortrainierte Modelle als Ausgangspunkt verwendet. Mit keinem der genann-

ten Modelle konnte der gesuchte Zusammenhang zwischen Schnittkante und Prozessparametern abgebildet werden. Deshalb wurde eine eigene CNN-Architektur entwickelt, die vom VGG16-Netz inspiriert wurde, aber deutlich weniger zu erlernende Parameter hat und die deshalb mit der zur Verfügung stehenden Datenbasis trainiert werden konnte.

5.1.1.1 VGG16-Netz

Das VGG16-Netz wurde 2014 von Simonyan und Zisserman [SZ15] vorgestellt. Die Autoren haben mit dieser Architektur den ersten und zweiten Platz der ImageNet Challenge 2014 in den Kategorien Lokalisierung und Klassifizierung belegt. Bei dem Wettbewerb ging es darum, ein Modell zu trainieren, das 1000 verschiedene Bildkategorien der Datenbank ImageNet [Den+09] erkennen und die Objekte im Bild lokalisieren kann. Seitdem wurde immer wieder gezeigt, dass die Architektur auch gut für andere Daten funktioniert (z. B. zur Detektion von Kiwis auf Basis von RGB- und Nahinfrarot-Bilddaten [Liu+20]).

Das VGG16-Netz besteht aus 16 Schichten, die trainierbare Parameter enthalten. Davon sind 13 Faltungsschichten und daran anschließend drei vollständig verbundene Schichten. Die Faltungsschichten sind in Blöcken organisiert: Die ersten beiden Blöcke enthalten jeweils zwei, die folgenden drei jeweils drei Faltungsschichten. Zentrale Ideen des VGG16-Netzes sind die Verwendung von 3×3 großen Faltungskernen und die mit fortschreitender Tiefe immer stärkere Reduktion der Größe der Feature Maps. Die Größe halbiert sich nach jedem Block von anfangs 224×224 Pixeln auf 7×7 Pixel nach dem fünften Block. Die Tiefe der Schichten steigt vom Eingang zum Ausgang hin an. Während der erste Block 64 Feature Maps aufweist, sind es im zweiten Block 128, im dritten 256 und im vierten und fünften Block sogar 512.

Mit 138 Millionen hat das VGG16-Netz zu viele Parameter, um mit der verfügbaren Datenbasis (siehe Abschnitt 4.5) von Grund auf trainiert zu werden. Deshalb wurde getestet, ob eine auf den 14 Millionen Bildern der ImageNet-Datenbank vortrainierte Version des Modells verwendet werden kann. Dazu wurden die Parameter der Faltungsschichten, welche die Merkmalsextraktion durchführen, unverändert übernommen und nur die Parameter der vollständig verbundenen Schichten, welche die Zuordnung zwischen Merkmalen und Kategorien vornehmen, wurden neu initiali-

Tabelle 5.1 Architektur des CNNs zur Regression der Prozessparameter aus einem Bild der Schnittkante. B bezeichnet die Nummer des Blocks: $B = \{1, 2, 3, 4\}$. Nicht eingetragen sind die in jeder Schicht (außer GAP, Dropout) angewendete Aktivierungsfunktion sowie die Batch-Normalisierung.

Schicht	Eingangsgröße	Tiefe	Ausgangsgröße
Eingangsblock	224×224	32	112×112
B , Conv I	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^B \times 224/2^B$
B , Conv II	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^B \times 224/2^B$
B , Conv III	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^{B+1} \times 224/2^{B+1}$
GAP	7×7	–	256
Dropout	256	–	256
FC	256	–	5

siert und während des Trainings optimiert. Es zeigte sich jedoch, dass die vortrainierten Schichten keine sinnvollen Merkmale aus den Schnittkantenbildern extrahieren konnten. Das ist nicht verwunderlich, da die ImageNet-Datenbank zwar Texturbilder enthält, aber im Wesentlichen Bilder alltäglicher Objekte beinhaltet, deren Merkmale und Strukturen sich erheblich von denen von Laserschnittkanten unterscheiden. Deshalb wurde eine eigene CNN-Architektur entwickelt und von Grund auf trainiert.

5.1.1.2 CNN-Architektur

Die entwickelte CNN-Architektur orientiert sich an den zentralen Ideen des VGG-Netzes, hat aber mit 4 statt 138 Millionen eine deutlich kleinere Parameteranzahl. Des Weiteren wurden Forschungsergebnisse einbezogen, die nach der Veröffentlichung des VGG-Netzes vorgestellt wurden. Die Netzarchitektur ist in Tabelle 5.1 zusammengefasst. Eine vereinfachte Darstellung ist in Abbildung 5.1 gezeigt.

Am Eingang des Netzes wird ein Bild mit den Abmessungen 224×224 Pixel und mit einem, drei oder vier Kanälen übergeben. Am Ausgang werden die Werte von fünf Prozessparametern (ADB, EM, FP, Gasdruck, Vorschub) geschätzt. Das CNN besteht aus 13 Faltungsschichten und einer vollständig verbundenen Schicht, die die letzte verdeckte Schicht mit den

fünf Ausgängen verbindet. Die Schichten sind in Blöcken organisiert. Es gibt einen Eingangsblock bestehend aus einer Faltungsschicht. Dieser wird von der von He et al. [He+16] vorgestellten ResNet-Architektur übernommen. Dort werden 32 Filter mit einer Größe von 7×7 und einer Schrittweite von 2 verwendet. Diese Filter haben gegenüber den ansonsten genutzten 3×3 großen Filtern einen größeren rezeptiven Bereich, was sich vorteilhaft auswirken kann. Die Größe des Eingangsbildes wird durch den Eingangsblock von 224×224 auf 112×112 Pixel verringert.

An den Eingangsblock angeschlossen sind vier Blöcke mit jeweils drei Faltungsschichten: Conv I, II und III. Die Filter haben alle die Dimension 3×3 . Während sich die Anzahl der Filter mit jedem Block verdoppelt, von 32 bis 256, reduziert sich die räumliche Größe des Bildes von anfangs 224×224 auf 7×7 Pixel. Anders als beim VGG16-Netz wird dafür kein Max Pooling verwendet, sondern die Conv III-Faltungsschichten, deren Filter mit einer Schrittweite von 2 (statt wie sonst 1) über die Feature Maps der Vorgängerschicht geschoben werden. Diese Methode wurde von Springenberg et al. [Spr+15] vorgeschlagen und hat den Vorteil, dass das Pooling nicht einfach durch Übernahme eines Maximalwertes erfolgt, sondern parametrisiert werden kann.

Der Ausgang der letzten Faltungsschicht hat 256 Feature Maps mit den Dimensionen 7×7 . Im VGG16-Netz wird diese letzte Schicht vektorisiert und über drei vollständig verbundene Schichten mit den Ausgängen verbunden. Dies trägt maßgeblich zur großen Parameteranzahl bei. Hier wird Global Average Pooling in Kombination mit einer vollständig verbundenen Schicht verwendet. Es werden also nicht alle Neuronen der letzten Faltungsschicht mit dem Ausgang verbunden, sondern es wird der Durchschnitt jeder Feature Map verwendet. Dies führt zu 256 (statt 12544) Neuronen, die mit den fünf Ausgängen verbunden werden müssen. Um eine Überanpassung des Modells an die Trainingsdaten zu verhindern, wurden im Training mittels Dropout 50 % der Neuronen der GAP-Schicht ausgeschaltet.

Eine weitere Änderung gegenüber dem VGG16-Netz ist, dass nicht ReLU sondern PReLU als Aktivierungsfunktion verwendet wird. Dies erhöht die Parameteranzahl zwar, verkleinerte den Fehler auf dem Validierungsdatensatz aber deutlich.

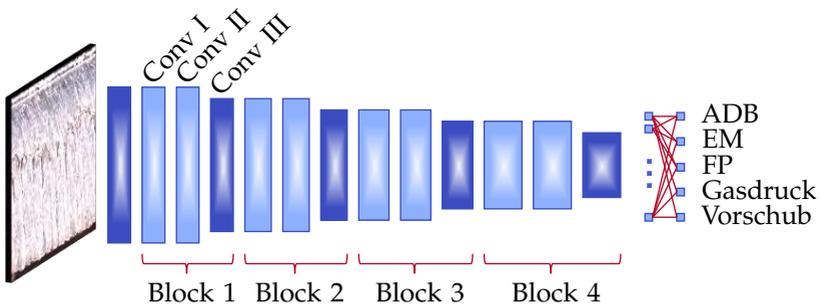


Abbildung 5.1 Architektur des CNNs zur Regression der Prozessparameter. Am Eingang wird ein Bild der Schnittkante übergeben. Das Netz besteht aus 13 Faltungsschichten, die in vier Blöcken organisiert sind. Die allererste Schicht und die letzte Schicht jedes Blocks (Conv III) reduzieren die räumliche Auflösung der Feature Maps jeweils auf ein Viertel. Sie sind dunkler dargestellt. Die Anzahl der Feature Maps pro Schicht nimmt mit jedem Block zu. Nach den Faltungsschichten bildet Global Average Pooling die 256 Feature Maps auf einen 256-dimensionalen Vektor ab. Dessen Neuronen werden mit den fünf Ausgängen verbunden.

Um eine bessere Vergleichbarkeit der fünf Regressionsziele zu erreichen und eine Über- oder Unterbewertung einzelner Prozessparameter zu vermeiden, werden diese jeweils auf den Bereich $[0, 1]$ skaliert. Während des Trainings wurde der mittlere quadratische Fehler (siehe Formel (3.6)) über alle Regressionsziele gemittelt als Verlustfunktion verwendet.

5.1.2 Klassifikation nach den Faktorstufen eines Prozessparameters

Für die Visualisierung der Zusammenhänge zwischen Schnittkante und Prozessparametern mittels LRP werden Klassifikatoren benötigt. Anders als bei der Regression, bei der mit einem Modell die Werte fünf verschiedener Prozessparameter geschätzt wurden, wird ein Modell pro Parameter trainiert. Es wird beispielsweise ein CNN trainiert, das ein Bild der Schnittkante einer der fünf Faktorstufen 13, 17, 21, 25 oder 29 m/min Vorschub zuordnet. Dafür wurde die in Tabelle 5.1 gezeigte Architektur wie folgt modifiziert: Um anstatt der Regression eine Klassifikation durchzuführen, wird eine Softmax-Funktion nach der letzten Schicht eingefügt. Diese bildet die Komponenten eines Vektors auf einen Vektor derselben Dimen-

Tabelle 5.2 Architektur der CNNs zur Klassifikation nach den Faktorstufen je eines Prozessparameters basierend auf einem Bild der Schnittkante. B ist die Nummer des Blocks: $B = \{1, 2, 3, 4\}$. Nicht eingetragen sind die in jeder Schicht (außer Pooling, GAP, Dropout) angewendete Aktivierungsfunktion sowie die Batch-Normalisierung. Diese Operationen können für die Visualisierung jedoch als Schichten interpretiert werden.

Schicht	Eingangsgröße	Tiefe	Ausgangsgröße
Eingangsblock	224×224	32	112×112
B , Conv I	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^B \times 224/2^B$
B , Conv II	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^B \times 224/2^B$
B , Max Pooling	$224/2^B \times 224/2^B$	$16 \cdot 2^B$	$224/2^{B+1} \times 224/2^{B+1}$
GAP	7×7 , Tiefe 256	–	256
Dropout	256	–	256
FC, Softmax	256	–	Anz. Klassen

sion ab, dessen Komponenten alle positiv sind und in Summe 1 ergeben [Bis06]. Die Ausgangswerte können deshalb als Wahrscheinlichkeiten der Zugehörigkeit zur jeweiligen Klasse interpretiert werden. Die PReLU-Aktivierungsfunktion wird durch ReLU ersetzt, weil die verwendeten Formeln für die LRP dies voraussetzen. Des Weiteren werden die Conv III-Faltungsschichten mit der Schrittweite 2 durch Max Pooling ersetzt. Diese Änderung dient nicht der Optimierung der Klassifikationsgüte. Die Architektur des CNNs wirkt sich aber auch auf das Aussehen der LRP-Heatmaps aus und diese waren bei der Verwendung von Max Pooling weniger verschwommen. Eine Zusammenfassung der Architektur ist in Tabelle 5.2 zu sehen.

Während des Trainings der Klassifikatoren wurde die kategorische Kreuzentropie (siehe Formel (3.9)) als Verlustfunktion verwendet und minimiert.

5.1.3 Schätzung der Rautiefe

Das CNN für die Schätzung der Rautiefe ist nahezu identisch zu dem in Tabelle 5.1 vorgestellten. Es hat jedoch neun (statt fünf) Ausgänge, einen für jede der neun Messlinien in verschiedenen Tiefen des Blechs. Die in den fünf Bereichen einer Messlinie ermittelten Rz -Werte (siehe

Abbildung 4.7) werden dafür gemittelt. Die 15 Bilder einer Kante werden mit denselben Rz -Werten gelabelt, da eine separate Messung der Rautiefe in jedem Bildbereich aufgrund der zu kleinen Messstrecke unmöglich ist.

Alle Neuronen verwenden ReLU- statt PReLU-Aktivierungsfunktionen, weil das die Anzahl zu trainierender Parameter und damit das Risiko einer Überanpassung an einzelne, fehlerhafte Trainingsdatenpunkte verringert. Während des Trainings wird der mittlere absolute Fehler (siehe Formel (3.7)) statt des mittleren quadratischen Fehlers minimiert. Bei dieser Metrik fallen einzelne große Abweichungen zwischen Vorhersage und Label weniger stark ins Gewicht.

5.1.4 Konfigurationen

Die hier beschriebenen Konfigurationen betreffen alle oben vorgestellten CNNs.

Die Optimierung der Modellparameter erfolgte mit dem Adam-Algorithmus [BK15], welcher eine Weiterentwicklung des Gradientenverfahrens ist. Es wurden die von den Autoren vorgeschlagenen Default-Parameter (Lernrate 0,001; β_1 0,9; β_2 0,999) verwendet.

Um die Speicheranforderungen während des Trainings möglichst gering zu halten, wurden die Trainingsbilder in Mini-Batches an das Modell übergeben. Ein Batch enthielt 60 Bilder. Batch Normalisierung wurde in jeder Faltungsschicht angewendet, um das Trainingstempo zu erhöhen.

Um Überanpassung zu vermeiden und die Robustheit der Modelle zu vergrößern, wurden die Trainingsbilder zufällig horizontal und vertikal gekippt und gedreht, bevor sie an das Netz übergeben wurden (engl. *online data augmentation*). Bei der Drehung der Bilder entstehende leere Pixel wurden mit Nullen, die schwarzer Farbe entsprechen, aufgefüllt. Aus demselben Grund wurde das Training beendet, sobald sich der Validierungsfehler über 20 Epochen (d. h. Iterationsschritte) nicht mehr verkleinerte (engl. *early stopping*). Verbesserte sich der Validierungsfehler beispielsweise zwischen Epoche 27 und 47 nicht weiter, wurde das Training nach 47 Epochen beendet und die Modellparameter der 27. Epoche gespeichert. Die Trainingszeit lag in allen Fällen unter zwölf Stunden und wird im Folgenden nicht weiter berücksichtigt.

Informationen zur verwendeten Hardware und zu den Versionen der eingesetzten Python-Bibliotheken sind in Anhang A.1 zu finden.

5.1.4.1 Aufteilung der Daten für Training, Validierung und Test

Die vorhandenen Daten, 50040 Bilder von Schnittkanten (siehe Tabelle 4.3), wurden in drei Teile geteilt, damit die CNNs trainiert und dann auf unbekanntem Bildern validiert und getestet werden konnten. Der Trainingsdatensatz beinhaltet 36060 Bilder. Für die Validierung wurden 6420 und für das Testen wurden 7560 Bilder verwendet.

Die 60 Bilder eines Teils (und damit einer Parameterkombination) sehen sehr ähnlich aus. Deshalb muss bei der Aufteilung in Trainings-, Validierungs- und Testdaten darauf geachtet werden, dass jede Parameterkombination nur in einem der drei Teile enthalten ist. Sonst könnte es sein, dass das CNN die Bilder aus dem Training während des Tests lediglich wiedererkennt und man die Modellgüte überschätzt.

Die Daten wurden dreimal zufällig unter Beachtung der oben beschriebenen Regel aufgeteilt. Die drei Aufteilungen werden im Folgenden mit Split 1, 2 bzw. 3 bezeichnet. Die zugehörigen Teilmengen der Daten heißen Trainings-, Validierungs- und Testset 1, 2 bzw. 3.

5.2 Visualisierung mit Layer-wise Relevance Propagation

In Abschnitt 3.2.6 befindet sich ein Überblick über die vielfältigen Möglichkeiten, komplexe ML-Modelle zu visualisieren und erklärbar zu machen. Die Grundidee der in dieser Arbeit genutzten Layer-wise Relevance Propagation ist es, die Klassifikationsentscheidung eines künstlichen neuronalen Netzes in die jeweiligen Beiträge (die sogenannten Relevanzen) der Neuronen der Eingangsebene zu zerlegen. Die so entstehenden Heatmaps haben dieselbe Dimension wie das Eingangsbild. Der Relevanzwert eines Pixels zeigt den Beitrag dieses Pixels zum Ausgangswert des Netzes an. Im Folgenden wird erklärt, wie die LRP-Regel(n) mithilfe der Deep Taylor Decomposition (DTD, dt. tiefe Taylor-Zerlegung) hergeleitet werden können. Anschließend werden Details zur Implementierung erläutert.

Vor der formalen Herleitung soll jedoch ein anschauliches Verständnis der Methode geschaffen werden. Dazu ist in Abbildung 5.2 ein vollständig verbundenes neuronales Netz dargestellt. Normalerweise (siehe dazu Abbildung 3.3) werden die Eingangsdaten, z. B. ein Bild, an die Neuronen

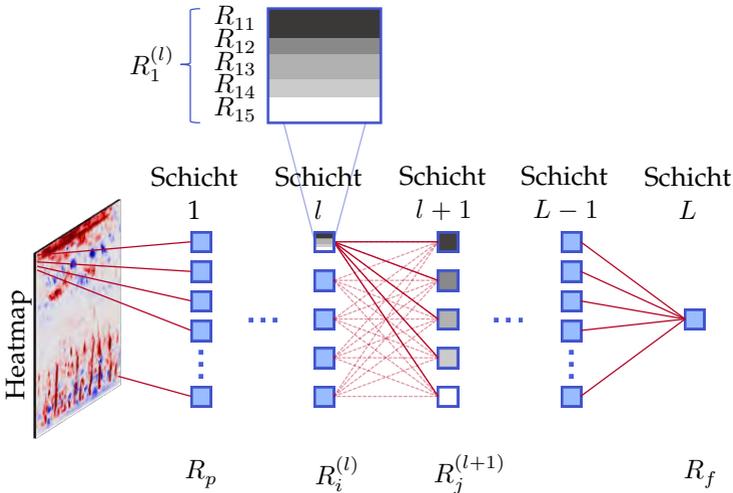


Abbildung 5.2 Berechnung des Beitrags jedes Pixels des Eingangsbildes zum Ausgangswert des neuronalen Netzes mit LRP. Der Ausgangswert wird Schicht für Schicht zurück zum Eingang propagiert. Das Ergebnis ist eine Heatmap mit derselben Dimension wie das Eingangsbild. Um die Relevanz eines Neurons (im Bild $R_{i=1}$) der Schicht l zu berechnen, müssen die Relevanzwerte aller Neurons der Schicht $l+1$ (im Bild $j = \{1, 2, 3, 4, 5\}$) gewichtet zurück propagiert (im Bild von rechts nach links) und anschließend summiert werden.

der ersten Schicht übergeben und dann wird die Information von Schicht zu Schicht weitergegeben bis zum Ausgang. Für die Visualisierung geht man anders herum vor. Die Aktivierung eines Ausgangsneurons, z. B. der Klasse 13 m/min, wird gewichtet an alle Neuronen der Vorgängerschicht übertragen und dann zurück zum Eingang propagiert. Zum Schluss werden die Aktivierungen auf die Pixel des Eingangsbildes aufgeteilt.

5.2.1 Deep Taylor Decomposition

Die Deep Taylor Decomposition ist eine Anwendung der Taylor-Reihe auf tiefe neuronale Netze. Ziel dieses Verfahrens ist es, den Funktionswert so auf die Eingangsvariablen zu verteilen, dass der jeweilige Anteil gerade ihrem Beitrag zum Funktionswert entspricht. Die folgende Herleitung ist, sofern nicht anders angegeben, Montavon et al. [Mon+17] entnommen.

Ein neuronales Netz (mit einem Ausgang) wird durch eine Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$ beschrieben. Das Funktionsargument $\mathbf{x} \in \mathbb{R}^d$ ist hier ein (vektorisertes) Bild mit P Pixelwerten x_p , wobei $p \in \{1, 2, \dots, P\}$. Handelt es sich bei dem neuronalen Netz um ein Modell, mit dem Bilder klassifiziert werden können, so gibt f die An- oder Abwesenheit eines bestimmten Merkmals im Bild an. $f(\mathbf{x}) = 0$ bedeutet, dass das Merkmal nicht vorhanden ist, $f(\mathbf{x}) > 0$ zeigt an, dass (und mit welcher Sicherheit) es vorhanden ist. Mit der DTD wird jedem Pixel ein Relevanzwert $R_p(\mathbf{x})$ zugewiesen. Die zu einem Eingangsbild gehörenden Relevanzwerte $\{R_p(\mathbf{x})\}$ werden als Heatmap bezeichnet. Diese Heatmap hat dieselbe Dimension wie das Eingangsbild. Sie soll die folgenden Eigenschaften haben:

Definition 5.1 (Erhaltend)

Eine Heatmap $\{R_p(\mathbf{x})\}$ ist erhaltend, wenn die Summe der Relevanzen $R_p(\mathbf{x})$ dem Ausgangswert des Modells R_f entspricht:
 $\forall \mathbf{x} : R_f = \sum_p R_p(\mathbf{x})$ wobei $R_f = f(\mathbf{x})$.

Definition 5.2 (Positiv)

Eine Heatmap $\{R_p(\mathbf{x})\}$ ist positiv, wenn alle ihre Werte gleich oder größer 0 sind:
 $\forall \mathbf{x}, p : R_p(\mathbf{x}) \geq 0$.

Eine Möglichkeit, den Funktionswert des neuronalen Netzes entsprechend den Definitionen (5.1) und (5.2) in die Beiträge der einzelnen Pixel zu zerlegen, stellt die Taylor-Reihenentwicklung dar:

$$R_f = f(\mathbf{x}) = f(\tilde{\mathbf{x}}) + \left(\frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}} \right)^\top \cdot (\mathbf{x} - \tilde{\mathbf{x}}) + \eta \quad (5.1)$$

$$= 0 + \sum_p \frac{\partial f}{\partial x_p} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}} \cdot (x_p - \tilde{x}_p) + \eta. \quad (5.2)$$

Die Reihenentwicklung erfolgt um die Entwicklungsstelle $\tilde{\mathbf{x}}$. Für diese soll gelten: $f(\tilde{\mathbf{x}}) = 0$. Terme zweiter und höherer Ordnung η werden nicht weiter berücksichtigt. Vergleicht man die Definition (5.1) und die obige

Gleichung (5.2), kann man $R_p(\mathbf{x})$ als den Term innerhalb des Summenzeichens identifizieren:

$$R_p(\mathbf{x}) := \left. \frac{\partial f}{\partial x_p} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} \cdot (x_p - \tilde{x}_p). \quad (5.3)$$

Für ein gegebenes neuronales Netz f stellt die Wahl der Entwicklungsstelle den einzigen Freiheitsgrad dieses Ansatzes dar. Für sie soll gelten, dass sie dem Datenpunkt \mathbf{x} (nach einer bestimmten Abstandsdefinition) möglichst ähnlich (also nah) sein sollte. Gleichzeitig darf sie gerade die Information nicht enthalten, die dazu führt, dass \mathbf{x} seiner Klasse zugeordnet wird, da gelten soll: $f(\tilde{\mathbf{x}}) = 0$. Anschaulich ist die Differenz zwischen \mathbf{x} und $\tilde{\mathbf{x}}$ dann ein Bild, welches nur noch die Klasseninformation enthält.

Bei der Berechnung des Gradienten und der Entwicklungsstelle in Formel (5.3) kann wie folgt vorgegangen werden. Ein KNN besteht aus mehreren Schichten, die jeweils mehrere Neuronen enthalten. Im Folgenden wird – der Einfachheit wegen – von vollständig verbundenen Schichten (siehe Abschnitt 3.2.1.1) ausgegangen: Alle Neuronen $\{a_i\}^{(l)}$ einer Schicht l sind über die Gewichte $w_{ij}^{(l,l+1)}$ mit allen Neuronen $\{a_j\}^{(l+1)}$ der darauf folgenden Schicht $l + 1$ verbunden. Für eine bessere Lesbarkeit werden die hochgestellten Indizes, die die Zugehörigkeit zur Schicht ausdrücken, im Folgenden meist weggelassen. Aktivierungen/Relevanzen/Modellparameter der Schicht l werden stets mit i , Aktivierungen/Relevanzen/Modellparameter der Schicht $l + 1$ werden mit j indiziert.

Während des Trainings werden die Modellparameter so angepasst, dass der Fehler der Vorhersage am Ausgang auf dem Trainingsdatenset minimiert wird. Dabei wird eine hierarchische Struktur erlernt, z. B. reagiert jedes Neuron der ersten Schicht auf eine bestimmte, einfache Struktur im Eingangsbild. In tieferen Schichten werden diese einfachen zu immer komplexeren Strukturen zusammengesetzt [Arr+16].

Die DTD-Methode ist von diesem hierarchischen Prinzip inspiriert. Statt das neuronale Netz als Ganzes zu verstehen, betrachtet man wie in Formel (3.11) nur die lokale Beziehung der Neuronen $\{a_i\}$ einer Schicht zu einem Neuron a_j in der darauffolgenden Schicht.

Unter der Annahme, dass diese beiden Objekte durch eine lokale Funktion $R_j(\{a_i\})$ verbunden sind und dass es eine Entwicklungsstelle gibt, für die $R_j(\{\tilde{a}_i\}) = 0$ gilt, kann die Taylor-Reihenentwicklung darauf ange-

wendet werden, um die Beiträge $\{R_{ij}\}$ der Neuronen der Schicht l zum Ausgang a_j zu ermitteln:

$$a_j = R_j(\{a_i\}) \quad (5.4)$$

$$= \left(\frac{\partial R_j}{\partial \{a_i\}} \Big|_{\{a_i\}=\{\tilde{a}_i\}^{(j)}} \right)^\top \cdot (\{a_i\} - \{\tilde{a}_i\}^{(j)}) + \eta_j \quad (5.5)$$

$$= \sum_i \underbrace{\frac{\partial R_j}{\partial a_i} \Big|_{\{a_i\}=\{\tilde{a}_i\}^{(j)}}}_{:=R_{ij}} \cdot (a_i - \tilde{a}_i^{(j)}) + \eta_j. \quad (5.6)$$

Je größer R_{ij} desto wichtiger ist der Beitrag des Neurons i zum Neuron j . Es ist zu beachten, dass für jedes Neuron j eine DTD mit einer eigenen Entwicklungsstelle $\{\tilde{a}_i\}^{(j)}$ durchgeführt wird.

Die Rückpropagation der Relevanz beginnt am Ausgang des Netzes mit der Berechnung der Beiträge $\{R_{ij}\}$ der Neuronen der vorletzten Schicht ($L-1$) zu einem Neuron der Ausgangsschicht (L). Um die Relevanz Schicht für Schicht durch das Netz zurück zu propagieren, muss die gesamte Relevanz R_i eines Neurons der Schicht $L-1$ bestimmt werden, damit darauf wieder die Taylor-Reihenentwicklung bezüglich der Schicht $L-2$ angewendet werden kann. Dazu werden die Relevanzen R_{ij} aufsummiert:

$$R_i = \sum_j R_{ij}. \quad (5.7)$$

Aus der Kombination der letzten beiden Formeln ergibt sich

$$R_i = \sum_j \frac{\partial R_j}{\partial a_i} \Big|_{\{a_i\}=\{\tilde{a}_i\}^{(j)}} \cdot (a_i - \tilde{a}_i^{(j)}). \quad (5.8)$$

Für die Berechnung der Relevanz R_i des Neurons i , also seines Beitrags zu allen Neuronen der darauffolgenden Schicht, muss j -mal eine DTD durchgeführt und die Ergebnisse müssen addiert werden. Je nachdem welche Bedingungen für die Entwicklungsstelle gelten, können aus 5.8 verschiedene Regeln für die Rückpropagation der Relevanz abgeleitet werden.

In einem vollständig verbundenen neuronalen Netz mit ReLU-Aktivierung ist jedes Neuron mit der Vorgängerschicht über

$$a_j = \max \left(0, \sum_i a_i w_{ij} + b_j \right) \quad (5.9)$$

verbunden. Im einfachsten Fall, wenn jede reelle Zahl ein möglicher Eingangswert und $R_j > 0$ ist, entspricht die Entwicklungsstelle $\{\tilde{a}_i\}^{(j)}$ dem Schnittpunkt der folgenden beiden Ebenen:

$$0 = \sum_i \tilde{a}_i^{(j)} w_{ij} + b_j \quad (5.10)$$

$$\{\tilde{a}_i\}^{(j)} = \{a_i\} + t \cdot \{v_i\}^{(j)}. \quad (5.11)$$

Die erste Gleichung folgt aus Formel (3.11) und der Bedingung, dass $R_j(\{\tilde{a}_i\}) = 0$ gelten soll. Die zweite beschreibt die Suche nach der Entwicklungsstelle im Eingangsraum des Neurons j . Gesucht wird in Richtung $\{v_i\}^{(j)}$ nach einem Punkt, der dem aktuellen Datenpunkt $\{a_i\}$ nach der euklidischen Abstandsdefinition möglichst nah ist. (Falls $R_j = 0$ ist, dann ist $\{a_i\}$ bereits eine Entwicklungsstelle.)

Bei einem unbeschränkten Eingangswertebereich gilt hier $v_i = w_{ij}$ und $t \in \mathbb{R}$. Das bedeutet, dass ausgehend von $\{a_i\}$ in Richtung von $\{w_{ij}\}$ (dem Gradienten von R_j) nach der Entwicklungsstelle gesucht werden kann. Das Auflösen dieser Gleichungen nach $\{\tilde{a}_i\}^{(j)}$ und Einsetzen des Ergebnisses in Gleichung (5.8) ergibt die LRP- w^2 -Regel

$$R_i = \sum_j \frac{w_{ij}^2}{\sum_i w_{ij}^2} \cdot R_j. \quad (5.12)$$

Diese erfüllt sowohl die Anforderung der Erhaltung (5.1) als auch die der Positivität (5.2). Um die Relevanzen der Eingangspixel R_p zu erhalten, wird das Netz – gerade umgekehrt wie sonst (Formel 3.11) – Schicht für Schicht vom Ausgang zum Eingang durchlaufen.

Eine detailliertere Herleitung der Formel (5.12) wurde von Montavon et al. [Mon+15] publiziert. Je nachdem welche Eingangswerte zugelassen sind und wie genau die Entwicklungsstelle gesucht wird, ergeben sich von Formel (5.12) abweichende Regeln. Des Weiteren können Modifikationen vorgenommen werden, die vorteilhafte numerische Eigenschaften

aufweisen oder zu einer leichter interpretierbaren (weil z. B. weniger veräuschten) Heatmap führen. Die Forderungen nach Erhaltung (5.1) und Positivität (5.2) müssen dafür teilweise gelockert werden. In dieser Arbeit wurden vier verschiedene Regeln kombiniert, wie es Montavon et al. [Mon+19] vorgeschlagen hatten. Dadurch konnte eine möglichst klare und gleichzeitig verlässliche Darstellung erreicht werden. Die vier Regeln und ihre Kombination sind in Anhang A.4 erläutert.

5.2.2 Implementierungsdetails

Bislang wurde die LRP nur für vollständig verbundene Schichten hergeleitet. In dem CNN, dessen Entscheidungen in dieser Arbeit mithilfe der LRP visualisiert werden (siehe Abschnitt 5.1.2), werden jedoch verschiedene Operationen durchgeführt, die jede für sich als eine Schicht mit einer bestimmten Abbildungsvorschrift interpretiert werden können: vollständig verbundene Schicht, Faltungsschichten, Pooling (Global Average Pooling, Max Pooling), ReLU-Aktivierungsfunktion, Batch-Normalisierung und Softmax. Sie lassen sich alle auf die lineare Abbildung

$$a_j^{(l+1)} = \sum_i a_i^{(l)} w_{ij}^{(l,l+1)} + b_j^{(l+1)} \quad (5.13)$$

oder auf eine komponentenweise monoton steigende, nicht lineare Abbildung

$$a_j^{(l+1)} = \text{act} \left(a_i^{(l)} \right) \quad (5.14)$$

zurückführen [Lap19].

Vollständig verbundene Schichten und Faltungsschichten Vollständig verbundene Schichten und Faltungsschichten führen eine lineare Abbildung durch. Bei den Faltungsschichten sind dabei nicht alle Neuronen der Schicht l mit allen Neuronen der Schicht $l + 1$ verbunden. Für einen Großteil der Gewichte gilt $w_{ij} = 0$. $w_{ij} \neq 0$ sind die Parameter der Faltungskerne. Die Adaption der LRP-Regeln auf diese Schichten ist trivial.

Pooling-Schichten Bei den Pooling-Schichten wird zwischen Average und Max Pooling unterschieden. Ersteres kann als lineare Abbildung mit den Gewichten $w_{ij} = 1/N$ interpretiert werden, wobei N der Anzahl der zusammengefassten Neuronen entspricht. Bei der Vorwärts-Propagation der Aktivierungen wird beim Max Pooling nur die größte Aktivierung an die nächste Schicht weitergegeben. Analog dazu könnte die gesamte Relevanz bei der Rückpropagation an das maximal aktivierte Neuron übertragen werden. In dieser Arbeit wurde jedoch, wie von Montavon et al. [Mon+19] vorgeschlagen, die Relevanz wie beim Average Pooling auf alle beteiligten Neuronen verteilt.

Aktivierungsfunktion Bei der ReLU-Aktivierungsfunktion handelt es sich um eine nicht lineare Abbildung, die komponentenweise durchgeführt wird. Es werden keine Aktivierungen verschiedener Neuronen kombiniert. In diesem Fall gilt: $R_i^{(l)} = R_j^{(l+1)}$ mit $i = j$.

Batch-Normalisierung In jeder BN-Schicht findet eine komponentenweise Standardisierung der Aktivierungen statt (siehe Formel (3.16)). Die Batch-Normalisierung kann in drei Teilschritte zerlegt werden: Subtraktion von μ_{Batch} , Skalierung mit $\chi\sigma_{\text{Batch}}^{-1}$ und Addition von ϕ . Daraus ergibt sich [Lap19]:

$$R_i^{(l)} = \frac{a_i^{(l)} \cdot (\chi\sigma_{\text{Batch}}^{-1}) \cdot R_j^{(l+1)}}{a_j^{(l+1)}} \text{ mit } i = j. \quad (5.15)$$

Softmax-Funktion Die Softmax-Funktion wurde bei der LRP, der Empfehlung von Simonyan et al. [SVZ14] folgend, weggelassen.

Informationen zu den verwendeten Python-Bibliotheken sind in Anhang A.1 zu finden.

5.2.2.1 Visualisierung der Heatmaps

Das Ergebnis der LRP ist ein Relevanzwert für jedes Pixel und jeden Kanal des Eingangsbildes. Die Gesamtrelevanz eines Pixels ergibt sich aus der Addition der Relevanzen der einzelnen Kanäle. Die resultierenden

$P = 224 \cdot 224$ Relevanzwerte $\{R_p\}$ werden für die graphische Darstellung als Heatmap wie folgt transformiert:

$$\hat{R}_p = \frac{\frac{R_p}{\psi} + 1}{2} \text{ mit } \psi = 10 \cdot \left(\frac{1}{P} \sum_p |R_p|^3 \right)^{\frac{1}{3}}. \quad (5.16)$$

Dadurch werden negative Relevanzwerte auf $\hat{R}_p < 0,5$, positive Relevanzwerte auf $\hat{R}_p > 0,5$ abgebildet. ψ dient lediglich der Skalierung für eine möglichst „schöne“ Darstellung. Deshalb werden negative \hat{R}_p -Werte auch auf 0, \hat{R}_p -Werte größer als 1 werden auf 1 gesetzt. Zu beachten ist, dass durch dieses Abschneiden vom \hat{R}_p -Wert eines Pixels nur noch eingeschränkt auf seinen Anteil am Ausgangswert geschlossen werden kann.

\hat{R}_p -Werte kleiner als 0,5 werden blau eingefärbt und entsprechen einem negativen Beitrag des Pixels ($R_p < 0$), \hat{R}_p -Werte zwischen 0,5 und 1 werden rot eingefärbt und beschreiben einen positiven Beitrag des Pixels ($R_p > 0$) zum Ausgangswert des Netzes. Negativ relevante Pixel beinhalten Bildmerkmale, die nicht zu der vom Netz angegebenen Klasse passen. Dass es überhaupt negative Relevanzwerte gibt, hat mit der Lockerung der Positivitäts-Bedingung bei manchen der verwendeten LRP-Regeln zugunsten anderer Eigenschaften zu tun (für Details siehe Anhang A.4).

5.3 Zusammenhang zwischen Parametern und Kantenqualität

Für die Modellierung des Zusammenhangs zwischen den Prozessparametern und dem Fehlschnitttrisiko, der Rautiefe und der Grathöhe wurden einfache Modellansätze gewählt. Das Fehlschnitttrisiko der Faktorstufe eines Prozessparameters ist der Quotient aus der Anzahl der Fehlschnitte geteilt durch die Summe aller Schnitte. Das Fehlschnitttrisiko wurde mit der in Formel (3.4) eingeführten logistischen Regression modelliert, wobei 1 einem Fehlschnitt und 0 einem erfolgreichen Schnitt entspricht. Für die Rautiefe und die Grathöhe wurde der lineare Regressionsansatz (3.3) gewählt.

Die unabhängigen Variablen waren die vier Prozessparameter ADB, Fokusposition, Gasdruck und Vorschub. Damit nicht nur lineare Zusammen-

hänge abgebildet werden können, wurden diese vier Größen polynomial erweitert. Dabei entstanden alle Kombinationen der Prozessparameter kleiner und gleich dem angegebenen Grad. Bei zwei Eingangsvariablen $[x_1, x_2]$ würde die Erweiterung zweiten Grades beispielsweise die Variablen $[1, x_1, x_2, x_1x_2, x_1^2, x_2^2]$ ergeben. Nach der polynomialen Erweiterung wurden die entstehenden Variablen normiert, um eine Über- oder Unterbewertung einzelner Einflussfaktoren zu vermeiden. Außerdem wurden die Kostenfunktionen durch einen Regularisierungsterm $\|w\|_2^2$ ergänzt, der verhindern sollte, dass einzelne Gewichte zu groß werden.

Als Label dienten die Klassenzuordnung Fehlschnitt oder erfolgreicher Schnitt beziehungsweise die Rautiefen- und Gratwerte. Für die Rautiefe wurden vier Modelle trainiert: Rautiefe der Messlinien 03, 15 bzw. 27 und gemittelte Rautiefe. Zu jeder Parameterkombination lagen (wegen der vier Kanten) mehrere Werte für die Rautiefe und die Grathöhe vor. Die zu einer Parameterkombination gehörenden Datenpunkte wurden bei der Aufteilung in Trainings-, Validierungs- und Testdaten stets gemeinsam einer Teilmenge zugeordnet. Um eine zuverlässigere Aussage über die Modellgüte zu erhalten, wurde vierfach Kreuzvalidiert.

Die Modellgüte wurde über Precision und Recall (siehe Abschnitt 3.1.2) bzw. den mittleren absoluten Fehler (3.7) beschrieben. Sie wurde jeweils für alle Erweiterungen vom Grad 1 bis 7 verglichen. Für die Visualisierung der Zusammenhänge wurde je eine polynomiale Erweiterung ausgewählt.

Zu beachten ist, dass die Datenbasis nicht perfekt ausgeglichen war. Ein Ungleichgewicht ergab sich erstens durch die Erweiterung der vier vollfaktoriell variierten Prozessparameter durch die Fokusposition (siehe Abbildung 4.1(b)). Zweitens veränderten die Fehlschnitte, die nicht für alle Faktorstufen gleich wahrscheinlich sind (siehe Abbildung 6.25), die Verteilung der Datenpunkte für die Modellierung der Rautiefe und der Grathöhe. Drittens kam die LTT-Parameterkombination aufgrund der Wiederholung bei jedem 50. Teil besonders oft vor. Um dies zu verhindern, wurden alle LTT-Teile bis auf eines aus der Datenbasis entfernt.

Informationen zu den für die Implementierung genutzten Python-Bibliotheken sind in Anhang A.1 zu finden. Alle in diesem Kapitel vorgestellten Modellansätze sind in Tabelle 5.3 zusammengefasst. Die Grundlage für die Visualisierung mittels LRP bilden die CNN-Klassifikatoren 5.1.2.

Tabelle 5.3 Übersicht über die Modellansätze.

Eingangsdaten	Ausgangsdaten	Modelltyp
RGB-Bild	Prozessparameter	CNN 5.1.1.2
Höhenbild	Prozessparameter	CNN 5.1.1.2
RGBD-Bild	Prozessparameter	CNN 5.1.1.2
RGB-Bild	Faktorstufen ADB	CNN 5.1.2
RGB-Bild	Faktorstufen EM	CNN 5.1.2
RGB-Bild	Faktorstufen FP	CNN 5.1.2
RGB-Bild	Faktorstufen Gasdruck	CNN 5.1.2
RGB-Bild	Faktorstufen Vorschub	CNN 5.1.2
RGB-Bild	Rautiefe M_L 03 bis 27	CNN 5.1.3
Prozessparameter	Fehlschnitttrisiko	Log. Regression 5.3
Prozessparameter	Rautiefe	Poly. Regression 5.3
Prozessparameter	Grathöhe	Poly. Regression 5.3

6 Ergebnisse und Diskussion

Das Kapitel ist in vier Themenblöcke gegliedert. Zunächst wird die Datenbasis vorgestellt und analysiert. Anschließend werden zum Thema Schnittkantenqualität die Ergebnisse der Expertenbefragung sowie des CNNs zur bildbasierten Schätzung der Rautiefe präsentiert und diskutiert. Der dritte Teil behandelt die automatisierte Überprüfung des Maschinenzustands und die bildbasierte Regression der Prozessparameter. Abschließend soll das Verständnis des Laserschneidprozesses erweitert werden. Es wird aufgezeigt, wie die Visualisierung der von einem CNN erlernten Muster das Prozessverständnis erweitern und in der Praxis unterstützen kann. Daran anschließend werden die Zusammenhänge zwischen den Prozessparametern und dem Fehlschnittisiko, der Rautiefe und der Grathöhe analysiert und modelliert.

6.1 Datenbasis

In diesem Abschnitt wird die Datenbasis vorgestellt. Ein besonderes Augenmerk wird auf die Datenqualität gelegt und es werden verschiedene Ansätze zur Verbesserung derselben evaluiert.

6.1.1 Fehlschnitte

Der Versuchsplan zur Generierung der Datenbasis bestand aus 1126 Parameterkombinationen, wovon 1050 verschieden waren. Bei den restlichen 76 handelte es sich um zufällige Wiederholungen und um die LTT-Parameterkombination (siehe Abschnitt 4.1). 227 der 1126 Teile, also ca. 20 %, waren Fehlschnitte. Darunter waren 220 verschiedene Parameterkombinationen. Sie konnten nicht aus dem Restgitter entnommen und deshalb nicht weiterverwendet werden. Allerdings kann diese Information

Tabelle 6.1 Parameterkombinationen, die bei einem Teil zu einem Fehlschnitt führten und bei einem anderen Teil das Blech vollständig trennten mit Angabe der Teilenummer bei der ein/kein Fehlschnitt (FS) auftrat.

ADB in mm	EM in mm	FP in mm	Gasdruck in bar	Vorschub in m/min	FS	kein FS
0,5	-3,5	-3,0	9	17	0115	0144
2,0	-0,5	1,5	21	25	0135	0930
2,5	-1,5	1,0	15	29	0237	0556
3,0	-1,0	2,0	15	21	0916	0640

benutzt werden, um Rückschlüsse auf die Prozessstabilität in Abhängigkeit der Prozessparameter zu ziehen. Mehr zu diesem Thema befindet sich in Abschnitt 6.4.2.

In vier Fällen verursachte die gleiche Parameterkombination einmal einen Fehlschnitt und einmal trennte der Strahl das Blech vollständig. Diese vier Parameterkombinationen sind in Tabelle 6.1 aufgelistet. Die zweite und dritte Parameterkombination weisen einen gegenüber dem LTT-Wert erhöhten Vorschub und eine Fokusposition oberhalb des Blechs auf. Die erste und vierte Parameterkombination wurden zwar mit kleinen Vorschüben geschnitten, haben jedoch beide stark vom LTT-Wert abweichende Fokuspositionen.

Diskussion Bei einem hohen Vorschub und/oder bei einer zu hohen oder zu tiefen Position des Strahlfokus, wo die Energiedichte maximal ist, wird zu wenig Energie in das Blech eingebracht und es entsteht kein durchgehender Schnittspalt (mehr dazu in Abschnitt 6.4.2). Vermutlich befinden sich diese vier Parameterkombinationen genau an der Grenze des schneidbaren Prozessfensters. Deshalb entscheiden kleinste Schwankungen der Umgebungsbedingungen darüber, ob es zu einem Fehlschnitt kommt oder nicht.

Neben der Anzahl ist die Qualität der Daten für den Erfolg der datenbasierten Modellierung entscheidend. Diese wird in den folgenden Abschnitten für Rautiefe, Rechtwinkligkeitstoleranz und Grathöhe untersucht.

6.1.2 Rautiefe

In Abbildung 6.1 sind die Histogramme der Rautiefenwerte aller Messlinien und jeweils separat für die Messlinien 03, 15 und 27 gezeigt. Zu beachten ist, dass die Abszissen jeweils bei $80\ \mu\text{m}$ abgeschnitten wurden, es jedoch Werte gibt, die größer sind. Für die oberste Messlinie (03) beträgt der mittlere Rz -Wert $22\ \mu\text{m}$. Der Median befindet sich relativ weit davon entfernt bei $14\ \mu\text{m}$. Bei der mittleren (15) bzw. der untersten Messlinie (27) liegen Mittelwert und Median mit 18 und $15\ \mu\text{m}$ bzw. 23 und $21\ \mu\text{m}$ näher beieinander. In den meisten Fällen ist die Rautiefe im oberen Bereich der Schnittkante kleiner und nimmt dann über die Mitte nach unten hin zu. Gerade im oberen Bereich treten aber auch einige sehr große Rz -Werte auf.

Die Rautiefenwerte wurden von den Normvorgaben leicht abweichend, wie in Abschnitt 4.2.3.1 beschrieben, ermittelt. Für jede Kante erhält man $45\ Rz$ -Werte, fünf für jede der neun Messlinien in verschiedenen Tiefen des Blechs. Die fünf Rautiefenwerte derselben Messlinie auf derselben Kante sollten nahezu identisch sein, da für sie alle (bekannten) Einflussgrößen übereinstimmen. Diese Annahme wird nun für die Beurteilung der Datenqualität bei einmaligem Vermessen aller $3596 \cdot 5 = 17980$ Messwerten der Messlinie. Idealerweise sollten alle Messwerte eine Differenz von 0 aufweisen. Tatsächlich kommt es aber zu teilweise großen Abweichungen.

In Abbildung 6.2 ist die Verteilung der absoluten (4.3) und der relativen Differenz (4.4) separat für die Messlinien 03, 15 und 27 ausgewertet. Auf den Abszissen ist jeweils der Anteil der Messwerte in % aufgetragen, für den die Differenz kleiner oder gleich dem auf der zugehörigen Ordinate angegebenen Wert ist. Die 100% entsprechen allen $3596 \cdot 5 = 17980$ Messwerten der Messlinie. Idealerweise sollten alle Messwerte eine Differenz von 0 aufweisen. Tatsächlich kommt es aber zu teilweise großen Abweichungen.

Bei Messlinie 03 ist die mittlere absolute Differenz $8\ \mu\text{m}$, die mittlere relative Differenz ist $0,5$. 10% der Werte haben eine relative Differenz größer als 1 , sind also mehr als doppelt so groß wie der kleinste Wert in derselben Blechtiefe auf derselben Kante. Die mittlere absolute Differenz der Messlinie 15 ist $3\ \mu\text{m}$. Die entsprechende relative Differenz ist $0,2$. In 2% der Fälle ist die relative Differenz größer als 1 . Bei Messlinie 27 kommt es durchschnittlich zu Abweichungen von $3\ \mu\text{m}$, was einer relativen Differenz von ebenfalls $0,2$ entspricht. Nur 1% der Werte weisen eine relative Differenz größer als 1 auf. Sowohl die absoluten als auch die relativen Differenzen sind auf M_L 03 deutlich größer als auf M_L 15 und 27.

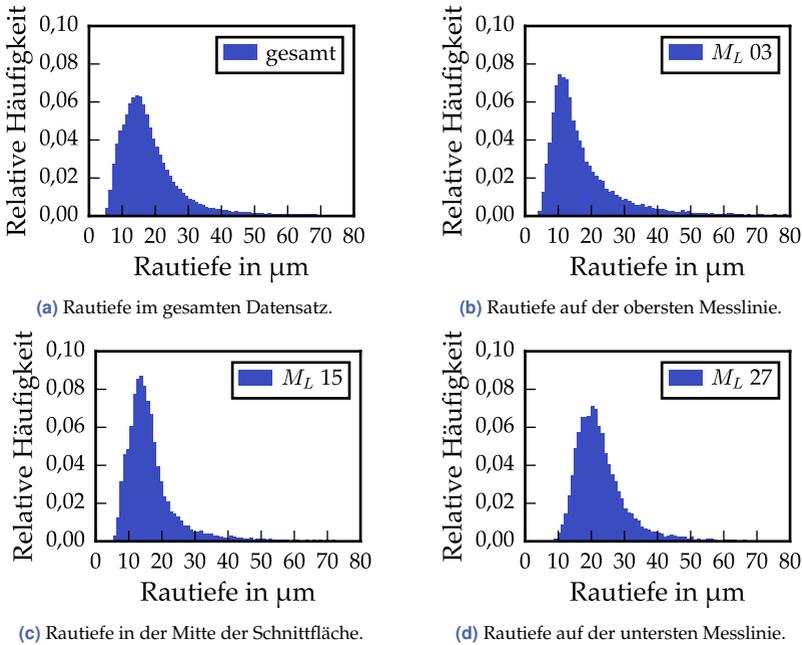
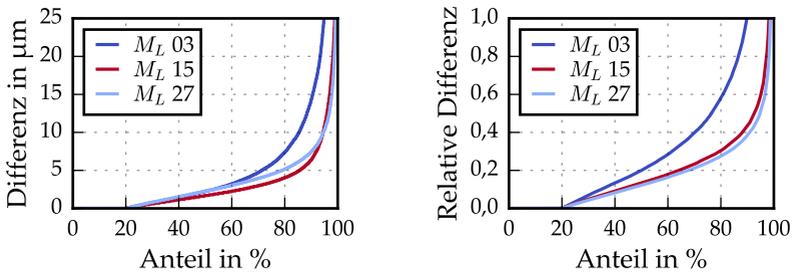


Abbildung 6.1 Verteilung der Rautiefenwerte des ersten Messdurchgangs für alle Messlinien sowie für die oberste (03), die mittlere (15) und die unterste (27) Messlinie.

Diskussion Vermutlich werden die Unterschiede der Rautiefenwerte durch die in Abschnitt 4.2.1.3 thematisierten Messfehler hervorgerufen. Ursächlich dafür könnte die hohe Reflektivität der Schnittkanten sein. Möglicherweise wird an stark reflektierenden Stellen nur wenig Licht in die Kamera gestreut und die Software des VR-3200 interpretiert dies fälschlicherweise als tiefe Senke in der Oberfläche. Eine andere Ursache für fehlerhafte Höhenwerte könnten Mehrfachreflexionen sein, die zu einer fehlerhaften lateralen Verschiebung (4.1) führen. Da der Rz -Wert auf der Höhendifferenz zwischen der tiefsten Senke und der höchsten Spitze basiert, führen tiefe Senken (und hohe Spitzen) zu sehr großen Rz -Werten.

In Abbildung 6.3 ist ein Beispiel gezeigt. Es sind jeweils die RGB-Bilder und die 3D-Punktwolken der Schnittfläche des ersten und eines zwei-



(a) Absolute Differenz der fünf Bereiche zum jeweils minimalen Rautiefenwert. (b) Relative Differenz der fünf Bereiche zum jeweils minimalen Rautiefenwert.

Abbildung 6.2 Verteilung der absoluten (4.3) und der relativen Differenz (4.4) der Rz -Werte des ersten Messdurchgangs.

ten Messdurchlaufs zu sehen. Die RGB-Bilder unterscheiden sich kaum, auf die 3D-Punktwolken trifft das nicht zu. Bei der ersten Messung sind tiefe Löcher (dunkelblau) im oberen Bereich der Schnittkante zu sehen, die bei der Wiederholung nicht auftreten. Die Rz -Werte auf Messlinie 03 der ersten Messung liegen zwischen 18 und 464 μm , die der zweiten Messung zwischen 18 und 23 μm , was angesichts der homogenen Schnittfläche deutlich plausibler ist. In diesem Beispiel ist der oberste Bereich der Schnittfläche besonders stark betroffen. Dass das für die gesamte Datenbasis gilt, bestätigt Abbildung 6.2, weil die Abweichungen auf Messlinie 03 besonders groß sind. Das könnte daran liegen, dass die Kanten im oberen Bereich meist sehr glatt sind und daher stärker reflektieren.

Diskussion Was die Identifikation der Messfehler schwierig macht ist, dass es neben den extremen Fällen (wie in Abbildung 6.3) sehr viele Kanten gibt, bei denen die Rautiefe durch Messfehler weniger stark verfälscht wird. Gleichzeitig ist die Bandbreite möglicher Rautiefenwerte im Datensatz groß, weswegen große Werte nicht zwangsläufig einen Fehler bedeuten. Auch innerhalb einer Schnittkante kann es zu Unterschieden kommen, wenn beispielsweise dicke Spritzer nur in der Mitte einer sonst glatten Kante auftreten. Die Reflexionsproblematik ließe sich möglicherweise durch die Verwendung eines anderen optischen Messverfahrens (z.B. konfokales Mikroskop), sicherlich durch die Verwendung eines nicht optischen Verfahrens verhindern. Ein solches steht jedoch nicht zur Verfügung

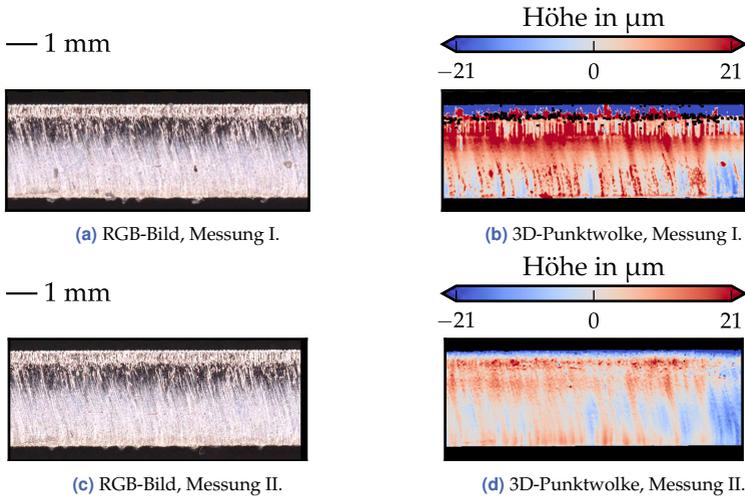


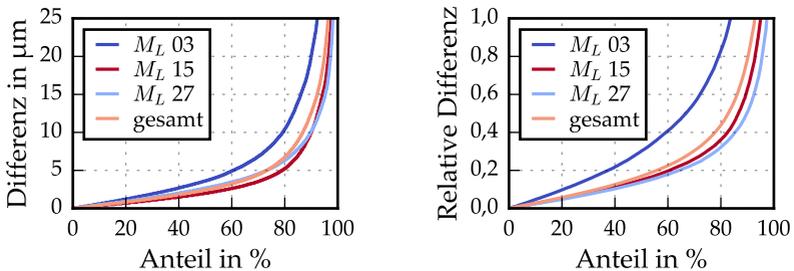
Abbildung 6.3 Messfehler durch hohe Reflektivität der Oberfläche, Ausschnitt von Kante 0703-1. Bei Messung I treten Fehler auf.

und wäre erheblich langsamer. Eine andere Option wäre die Beschichtung der Oberflächen z. B. mit Kreide. Dies kann jedoch ohne Voruntersuchungen, inwieweit eine solche Beschichtung die aufzunehmenden Riefen verschließt, nicht eingesetzt werden. Zweitens soll die Vermessung der Schnittkanten in Zukunft mit einem Roboter durchgeführt werden, für den das Aufbringen einer gleichmäßigen Kreidebeschichtung eine sehr schwierige Aufgabe darstellen würde. Möglicherweise wäre dem Problem durch mehrmaliges Einlegen und Messen beizukommen.

Um abschätzen zu können, wie viele Kanten von den Reflexionsproblemen betroffen sind, wurde jede Kante zweimal aufgenommen. Die Durchläufe werden mit I und II bezeichnet.

6.1.2.1 Quantifizierung des Messfehlers

In Abbildung 6.4 werden die gemessenen Rautiefenwerte von Messung I und II miteinander verglichen. Hier werden nicht die fünf Rz -Werte einer Messlinie und Kante zueinander in Beziehung gesetzt, sondern es werden



(a) Absolute Differenz zwischen Messung I und II. (b) Relative Differenz zwischen Messung I und II.

Abbildung 6.4 Verteilung der komponentenweisen absoluten und relativen Differenz der Rz -Werte zwischen Messdurchgang I und II.

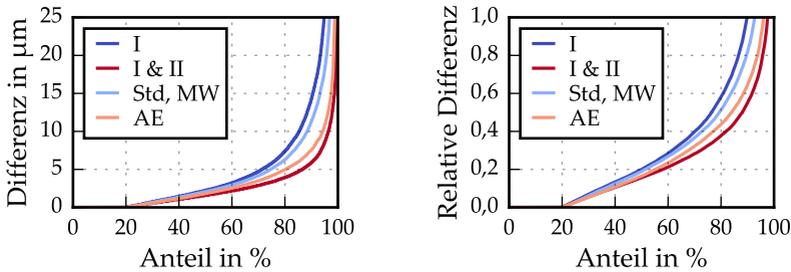
die Messergebnisse von Durchgang I und II komponentenweise (selbe Kante, gleiche Messlinie, gleicher Bereich) verglichen.

Für 93 % aller Messwerte ist die relative Differenz kleiner als 1. Das bedeutet umgekehrt, dass es in ca. 7 % der Fälle zu einem Unterschied zwischen Messung I und II kommt, der größer ist als der kleinere gemessene Wert. Ein Messwert, auf den das zutrifft, wird im Folgenden als Ausreißer eingestuft. Hier bestätigt sich die oben geäußerte Vermutung, dass im oberen Bereich mehr Messfehler auftreten als im mittleren und unteren Bereich der Schnittkante, da die Kurve für die Messlinie 03 stets deutlich über den anderen liegt. Die relative Differenz ist dort in 15 % der Fälle größer als 1.

Würde in beiden Messdurchgängen exakt derselbe Fehler auftreten, wäre die Differenz zwischen Messung I und II ebenfalls klein. Bei der manuellen Überprüfung einer Stichprobe wurde kein einziger solcher Fall entdeckt. Da die Messfehler die Datenqualität deutlich vermindern, werden im nächsten Abschnitt drei Methoden zur Korrektur der Fehler vorgestellt und bewertet.

6.1.2.2 Korrektur der Rautiefenwerte

In diesem Abschnitt werden drei Methoden zur Detektion und Korrektur der Messfehler verglichen. Diese wurden ausführlich in Abschnitt 4.3 beschrieben. Die Verteilung der absoluten und der relativen Differenz der fünf Rz -Werte auf Messlinie 03 ist in Abbildung 6.5 nach Anwendung



(a) Absolute Differenz der fünf Bereiche zum jeweils minimalen Rautiefenwert. (b) Relative Differenz der fünf Bereiche zum jeweils minimalen Rautiefenwert.

Abbildung 6.5 Vergleich der absoluten (4.3) und der relativen (4.4) Differenz der Rz -Werte auf Messlinie 03 ohne (Messdurchgang I) und mit einer Korrektur der Messfehler. I & II bezeichnet die Korrektur der Fehler durch Kombination der beiden Messdurchgänge. Std, MW beschreibt die Korrektur mit dem Mittelwert und AE die Korrektur mit dem Autoencoder. Mehr dazu steht in Abschnitt 4.3.

der verschiedenen Methoden gegenübergestellt. Die Verteilungen bei einmaligem Messen ohne weitere Korrektur, die in Abbildung 6.2 (auch in dunkelblau) gezeigt wurden, sind hier nochmal eingetragen. Die relative Differenz ist für 10 % der Bereiche größer als 1. Die mittlere absolute Differenz beträgt $7,7 \mu\text{m}$.

Kombination der Messungen I und II Die erste Methode zur Verbesserung der Datenqualität kombiniert die Ergebnisse von Messung I und II. Es wurde jeweils der kleinere Wert ausgewählt, weil – wie in den Abschnitten 4.2.1.3 und 4.2.3.1 erläutert – nur Fälle bekannt sind, in denen es durch Messfehler zu einer vergrößerten Rautiefe kommt. Durch die Kombination verbessert sich die Kurve hin zu kleineren Differenzen. Lediglich in 2 % (statt 10 %) der Fälle ist die relative Differenz größer als 1. Die mittlere absolute Differenz beträgt $2,5 \mu\text{m}$ (statt $7,7 \mu\text{m}$). Bei der Kombination werden 51 % der Datenpunkte von Messung I verändert, indem die Werte von Durchgang II eingesetzt werden. Die Wahrscheinlichkeit für Messfehler scheint in beiden Durchgängen nahezu gleich gewesen zu sein, da die Kombination die Ergebnisse von Durchgang I und II zu fast gleichen Teilen enthält. Dieser Ansatz ist mit einem erheblichen Mehraufwand verbunden, da alle Kanten zweimal vermessen werden müssen.

Die anderen beiden Methoden basieren nur auf den Messergebnissen des ersten Durchgangs und auf der Annahme, dass die *Rz*-Werte einer Messlinie über alle Bereiche und Kanten eines Teils ähnlich sind.

Korrektur über Std-Schwellenwert und Mittelwert Bei der Verwendung eines festen Schwellenwerts (Summe aus dem Mittelwert und dem 1,7-fachen der Standardabweichung der 20 *Rz*-Werte derselben Messlinie eines Teils) und dem Ersetzen aller Werte oberhalb dieser Schwelle mit dem neu berechneten Mittelwert reduziert sich der Anteil von Kanten mit einer relativen Differenz größer als 1 immerhin auf 7%. Die mittlere absolute Differenz reduziert sich von 7,7 auf 5,4 μm . Dabei werden 7% der Datenpunkte verändert.

Korrektur mit einem Autoencoder Die Korrektur mit einem Autoencoder geschieht in zwei Schritten. Zuerst werden alle 180 zu einem Teil verfügbaren Rautiefenwerte an den trainierten Autoencoder übergeben. Anschließend wird die Repräsentation am Ausgang komponentenweise mit dem Eingang verglichen. Ab einer relativen Differenz von 0,35 wird der Originalwert durch den vom Autoencoder berechneten Wert ersetzt. Nach dieser Korrektur tritt nur noch bei 4% (statt 10%) der Kanten eine relative Differenz größer als 1 auf. Die mittlere absolute Differenz beträgt 3,5 μm (statt 7,7 μm). Mit einem Schwellenwert von 0,35 in Formel (4.5) werden ebenfalls 7% der Datenpunkte verändert.

Diskussion Durch eine andere Wahl der Schwellenwerte kann der Anteil korrigierter Werte bei beiden Methoden verändert werden. Die Schwellen wurden jeweils so gewählt, dass ca. 7% der Datenpunkte verändert wurden, während 93% unverändert blieben. Diese Aufteilung wurde ausgehend von der Feststellung in Abschnitt 6.1.2.1, dass es in ca. 7% der Fälle zu Ausreißern kommt, festgelegt. Diese Ausreißerdefinition und die Veränderung von nur 7% der Werte ist ein vorsichtiger, aber bewusst gewählter Ansatz. Da die *Rz*-Werte als Label, also als grundlegende Wahrheit (engl. *ground truth*) für das Modelltraining dienen sollen, ist ihre Veränderung durch Algorithmen problematisch.

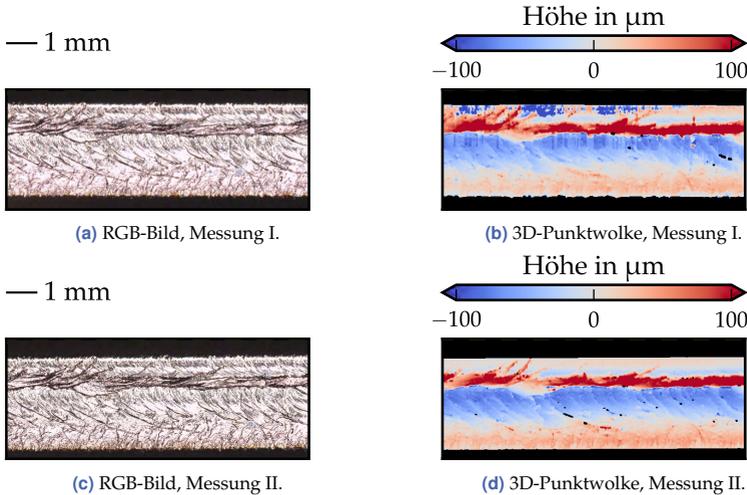


Abbildung 6.6 Messfehler durch hohe Reflektivität der Oberfläche, Ausschnitt von Kante 0584-4. Bei Messung I treten Fehler auf.

Obwohl beide Methoden (Std, MW und AE) gleich viele Datenpunkte verändern, erzielt der Autoencoder eine größere Verbesserung der Datenqualität. Warum die Korrektur mit dem Autoencoder vorteilhafter ist und bei welchen Kanten auch diese Methode nicht gut funktioniert, wird anhand der folgenden zwei Beispiele erklärt. In der Tabelle 6.2 aufgelistet sind die R_z -Messwerte der Kanten 0703-1 und 0584-4 auf Messlinie 03 für beide Messdurchgänge und die auf Messung I basierenden Korrekturmethode. Kante 0703-1 wurde bereits in Abbildung 6.3 gezeigt. Kante 0584-4 ist in Abbildung 6.6 zu sehen.

Beim ersten Messdurchgang weisen die R_z -Werte von Kante 0703-1 eine Differenz von $446 \mu\text{m}$ zwischen den Bereichen 1 und 4 auf. Beim zweiten Messdurchgang sind sowohl die R_z -Werte als auch die Differenzen zwischen den Bereichen deutlich kleiner. Die Messwerte der Kanten 2, 3 und 4 von Messung I sind ebenfalls aufgelistet. Der Mittelwert der R_z -Werte der vier Kanten beträgt $70 \mu\text{m}$, das 1,7-fache der Standardabweichung ist $206 \mu\text{m}$. R_z -Werte oberhalb von $276 \mu\text{m}$ gelten als Ausreißer. Die R_z -Werte 464 und $316 \mu\text{m}$ in den Bereichen 1 und 2 werden mit dem neu berechneten Mittelwert von $34 \mu\text{m}$ ersetzt, der R_z -Wert $101 \mu\text{m}$ in Bereich 3 nicht. Ein

Tabelle 6.2 *Rz*-Werte der Messlinie 03 (in μm) für zwei verschiedene Kanten, jeweils ohne und mit Messfehlerkorrektur.

Kante	Datensatz	M_B 1	M_B 2	M_B 3	M_B 4	M_B 5
0703-1	Messung I	464	316	101	18	28
	Messung II	23	17	18	18	21
	Std, MW	34	34	101	18	28
	AE	20	19	17	18	28
0703-2	Messung I	8	8	9	10	8
0703-3	Messung I	37	11	14	15	210
0703-4	Messung I	47	27	25	21	20
0584-4	Messung I	57	294	278	284	483
	Messung II	29	36	72	63	63
	Std, MW	57	294	278	284	183
	AE	57	158	278	284	284

potentiell falscher Wert wird also nicht als Ausreißer erkannt, außerdem ist der Korrekturwert von $34 \mu\text{m}$ vermutlich immer noch etwas zu groß.

Mit dem Autoencoder werden die *Rz*-Werte der Bereiche 1, 2 und 3 verkleinert. Ihre Größenordnung entspricht dann der des zweiten Messdurchgangs. Die *Rz*-Werte in Bereich 4 und 5 werden unverändert übernommen, weil die vom Autoencoder berechneten Werte nur geringfügig (weniger als vom Schwellenwert (4.5) vorgegeben) vom Original abweichen. Bei diesem Beispiel ist die Korrektur mit dem Autoencoder sehr erfolgreich. Anders sieht es bei Kante 0584-4 aus. Dort werden nur zwei der vermutlich vier falschen Rautiefenwerte mit etwas kleineren Werten ersetzt. Die Korrektur mit dem Autoencoder versagt bei sehr inhomogenen, aber größtenteils rauen Kanten in der direkten Umgebung dicker Spritzer, weil der Autoencoder anhand der Muster im näheren Umfeld auf die einzelnen Werte schließt.

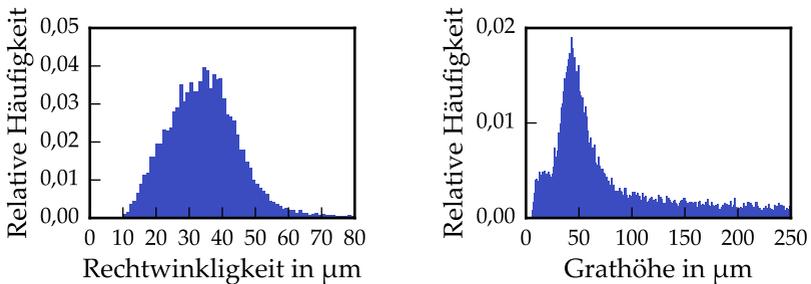
Diskussion Zusammenfassend kann gesagt werden, dass die Methode, die Messergebnisse mehrerer Durchgänge zu kombinieren, aus Gründen des Zeitaufwands nur bedingt empfehlenswert ist, obwohl sie die besten Ergebnisse liefert. Die Korrektur der *Rz*-Werte basierend auf dem Mit-

telwert und der Standardabweichung sollte nicht angewendet werden. Treten in mehreren Bereichen Fehler auf, führt dies zu einem hohen Mittelwert und zu einer großen Standardabweichung. Ausreißer werden nicht erkannt und falls doch, mit einem zu großen Wert ersetzt. Man könnte die Schwelle reduzieren, um mehr Datenpunkte als Ausreißer zu klassifizieren, aber dies würde einen größeren Anteil der originalen Daten verändern. Die Robustheit des Verfahrens zu erhöhen, indem nicht nur über die 20 *Rz*-Werte einer Messlinie, sondern über alle 180 Werte eines Teils gemittelt wird, ist nicht ratsam, da die Rautiefe in verschiedenen Tiefen des Blechs sehr unterschiedlich sein kann. Die verschiedenen horizontalen Schichten sind ein charakteristisches Merkmal von Laserschnittkanten [SM10], das nicht heraus gemittelt werden darf. Die Korrektur mit dem Autoencoder funktioniert besser. Hierfür werden nicht nur die 20 Werte einer Messlinie verwendet, sondern alle neun Messlinien. Der Autoencoder mittelt jedoch nicht einfach, sondern erlernt die typischen Muster der Schnittkanten und die Relationen der einzelnen Datenpunkte zueinander. Wie das Beispiel von Kante 0584-4 zeigt, funktioniert auch diese Methode nicht perfekt, aber sie reduziert die Messfehler deutlich und lässt dabei 93 % der Daten unverändert. Allerdings sollte ein zentraler Kritikpunkt des Autoencoders nicht unterschlagen werden: Es ist – im Gegensatz zu den anderen beiden Methoden – nicht klar, welche Muster und Beziehungen der Datenpunkte für die Rekonstruktion genutzt werden.

6.1.2.3 Zusammenfassung

Aus der aufgenommenen 3D-Punktwolke einer Schnittkante wurden 45 *Rz*-Werte extrahiert. Diese Werte sind fehlerbehaftet. Dies äußert sich in zu großen *Rz*-Werten. Besonders häufig ist dabei der oberste Bereich der Schnittkante betroffen. Eine zuverlässige Detektion dieses Fehlers ist sehr schwierig. Methodisch am einfachsten lässt sich die Datenqualität durch mehrmaliges Messen und die Kombination der Ergebnisse verbessern. Am zweitbesten funktioniert die Korrektur mit einem Autoencoder.

Die originalen bzw. korrigierten Rautiefenwerte der fünf Bereiche einer Messlinie werden für die weitere Verwendung gemittelt. Damit liegt pro Kante in jeder Blechtiefe genau ein Rautiefenwert vor. Im Folgenden werden – sofern nicht anders angegeben – die kombinierten Messdaten der Durchgänge I und II als beste vorhandene Datenbasis genutzt.



(a) Verteilung der Werte der Rechtwinkligkeitstoleranz.

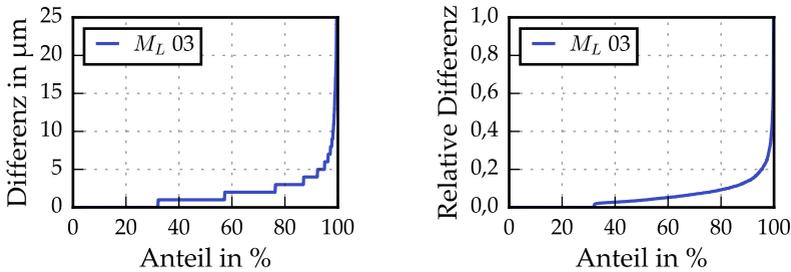
(b) Verteilung der Grathöhenwerte.

Abbildung 6.7 Verteilung der Werte von Rechtwinkligkeitstoleranz und Grathöhe im gesamten Datensatz des ersten Messdurchgangs.

6.1.3 Rechtwinkligkeitstoleranz

Die Rechtwinkligkeitstoleranz ist neben der Rautiefe die zweite in der Norm DIN EN ISO 9013 [Deu17] genannte Größe zur Beschreibung der Qualität einer Laserschnittkante. Sie wurde ebenfalls aus der 3D-Punktwolke der Schnittfläche abgeleitet und in fünf Bereichen gemessen (siehe Abschnitt 4.2.3.2). Die Messlinien verlaufen dabei orthogonal zu denen der Rautiefe. In Abbildung 6.7(a) ist ein Histogramm aller Werte des Datensatzes zu sehen. Der Mittelwert beträgt $35 \mu\text{m}$ und der Median ist $34 \mu\text{m}$, was bereits auf eine – verglichen mit der Rautiefe – geringere Ausreißerproblematik hindeutet. Abbildung 6.8 zeigt die Auswertung der Datenqualität – analog zur Rautiefe – über die Verteilung der absoluten und der relativen Differenz der fünf Messwerte einer Kante zum jeweiligen Minimum. Die Abweichung eines Messwertes zum kleinsten Wert derselben Kante ist durchschnittlich $2 \mu\text{m}$, das entspricht einer relativen Differenz von 0,07. Weniger als 1 % der Werte weisen eine relative Differenz größer als 1 auf.

Diskussion Die Datenqualität ist besser als bei der Rautiefenmessung, obwohl dieselben 3D-Punktwolken als Grundlage dienen. Da die fünf Messlinien aber kürzer sind und nicht entlang der Kante verlaufen, ist die Wahrscheinlichkeit, dass eine Messlinie einen Fehler beinhaltet, kleiner.



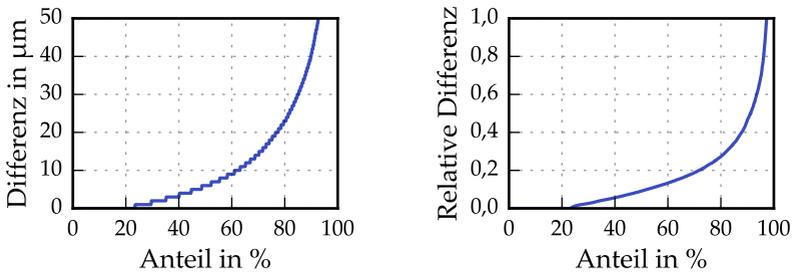
(a) Absolute Differenz der fünf Bereiche zum jeweils minimalen Rechtwinkligkeitswert. (b) Relative Differenz der fünf Bereiche zum jeweils minimalen Rechtwinkligkeitswert.

Abbildung 6.8 Verteilung der absoluten und der relativen Differenz (analog zu den Formeln (4.3) und (4.4)) der u -Werte für den ersten Messdurchgang.

6.1.4 Grathöhe

Die Grundlage für die Berechnung der Gratmerkmale bilden die aus der in Abbildung 4.4(d) gezeigten Perspektive aufgenommenen 3D-Punktwolken der Blechunterseite. In Abbildung 6.7(b) ist ein Histogramm der Grathöhenwerte im gesamten Datensatz gezeigt. Die durchschnittliche Grathöhe beträgt $104 \mu\text{m}$, der Median liegt mit $59 \mu\text{m}$ deutlich darunter. Da, anders als bei Rautiefe und Rechtwinkligkeitstoleranz, nicht pauschal davon ausgegangen werden kann, dass Messfehler zu vergrößerten Werten führen, kann das Minimum nicht als verlässlichster Wert gelten. Deshalb sind in Abbildung 6.9 die absolute und die relative Differenz der fünf Messwerte jeder Kante bezüglich des jeweiligen Medians aufgetragen. Die mittlere absolute Differenz ist $15 \mu\text{m}$, die mittlere relative Differenz beläuft sich auf $0,2$. In 3% der Bereiche ist die relative Differenz größer als 1.

Auch wenn ein direkter Vergleich mit der Messqualität der Rautiefenmessung schwierig ist, weil die Messwerte dort mit dem jeweils kleinsten und nicht mit dem Median verglichen werden, zeigt diese Auswertung, dass auch beim Grat Unterschiede entlang einer Kante auftreten. Um zu überprüfen, ob diese durch Messfehler hervorgerufen werden, werden die Messdurchgänge I und II im Folgenden miteinander verglichen.



(a) Absolute Differenz der fünf Bereiche zum jeweiligen Median. (b) Relative Differenz der fünf Bereiche zum jeweiligen Median.

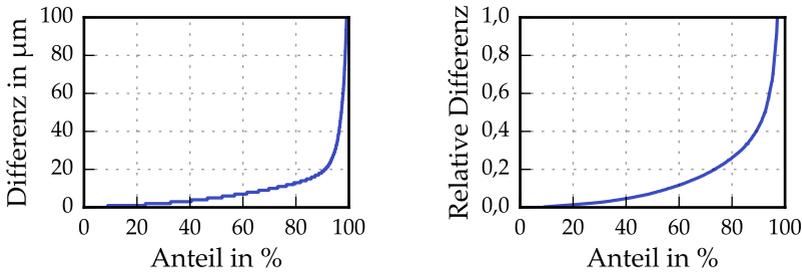
Abbildung 6.9 Verteilung der absoluten und der relativen Differenz der Grathöhen der Bereiche 1 bis 5 der jeweils selben Kante bezüglich des Medians.

6.1.4.1 Quantifizierung des Messfehlers

Analog zur Quantifizierung des Messfehlers bei der Rautiefe (siehe Abbildung 6.4) sind in Abbildung 6.10 die komponentenweise absolute und relative Differenz der Messwerte von Messdurchgang I und II gezeigt. In weniger als 3 % der Fälle unterscheiden sich die Grathöhen der beiden Messdurchgänge um mehr als das Doppelte des kleineren Werts. Das ist ein viel kleinerer Anteil als bei der Rautiefenmessung. Dort waren es z. B. 15 % auf Messlinie 03.

Eine stichprobenartige Überprüfung von Kanten mit großen Unterschieden ergab verschiedene Ursachen. Zum einen werden Abweichungen zwischen den beiden Messdurchgängen durch die Aufbewahrung und Handhabung der Proben verursacht, z. B. durch kleine Staubpartikel oder abgebrochene Gratperlen. Zum anderen treten auch hier sehr vereinzelt Messprobleme auf. Beispielsweise kommt es vor, dass der Grat aufgrund der geringen Tiefenschärfe von $\pm 0,5$ mm und der automatischen Fokussierung nicht korrekt abgebildet wurde.

Die Unterschiede der Grathöhe zwischen den verschiedenen Bereichen derselben Kante werden oft durch Prozessstreuungen und nur selten durch Messfehler hervorgerufen. Ein Beispiel dafür sind dicke Spritzer an der Unterseite der Kanten 2 oder 4. In der Laserschneidmaschine liegt das Blech auf Stegen auf (siehe Abschnitt 4.1.2). Befindet sich der Schnittspalt direkt über einer Stegspitze, kann das flüssig gewordene Material nicht



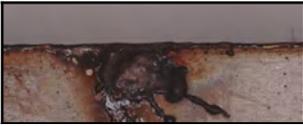
(a) Absolute Differenz zwischen Messung I und II. (b) Relative Differenz zwischen Messung I und II.

Abbildung 6.10 Verteilung der komponentenweisen absoluten und relativen Differenz der Grathöhenwerte zwischen Messdurchgang I und II.

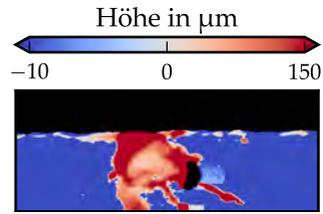
wie sonst aus dem Schnittpalt ausgetrieben werden. Dadurch werden, wie in den Abbildungen 6.11(a) und 6.11(b) gezeigt, sehr große Spritzer an der Blechunterseite hervorgerufen. Solche, durch Spritzer verfälschte Gratbereiche können mithilfe einer einfachen Ausreißerdetektion erkannt und entfernt werden, da an diesen Stellen sowohl Höhe als auch Breite des Grats um ein Vielfaches größer sind als in den anderen Bereichen. Zudem kann ein Spritzer in maximal zwei (benachbarten) Messbereichen auftreten, weil die Länge der vermessenen Kante kleiner ist als der Abstand zweier Stege. Große Differenzen zwischen den fünf Messbereichen treten außerdem bei Kanten auf, wie der in den Abbildungen 6.11(c) und 6.11(d) gezeigten. Der Grat dieser Kante besteht aus einzelnen Perlen, die sehr weit voneinander entfernt sind. Die Ursache dafür ist unklar.

Wenn man die Grathöhen der fünf Messbereiche zusammenfasst, liegt zu jeder Kante eine Grathöhe vor. Auffällig ist, dass es bei manchen Teilen zu erheblichen Unterschieden zwischen den vier Kanten desselben Teils kommt. Während auf einer Kante nahezu kein Grat vorhanden ist, sind auf einer anderen Kante dicke Tropfen an der Blechunterseite haften geblieben. Ein extremes Beispiel dafür ist Teil 0610. Die Grathöhe von Kante 2 (siehe Abbildung 6.11(e) und 6.11(f)) beträgt – gemittelt über die fünf Bereiche – 154 μm , die von Kante 4 (siehe Abbildung 6.11(g) und 6.11(h)) nur 10 μm . Diese beiden Kanten wurden mit denselben Parametern geschnitten. Es könnte also sein, dass es eine weitere Einflussgröße auf die Grathöhe gibt, die bislang nicht betrachtet wurde und die sich stärker auf den Grat als auf die Schnittfläche auswirkt, bei der solche Unterschiede nicht auftre-

– 1 mm



(a) Kante 0407-2, RGB-Bild.

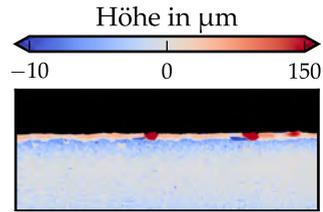


(b) Kante 0407-2, 3D-Punktwolke.

– 1 mm



(c) Kante 0446-2, RGB-Bild.

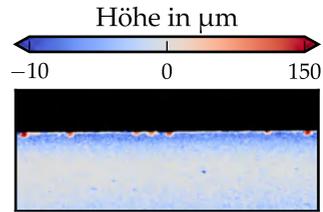


(d) Kante 0446-2, 3D-Punktwolke.

– 1 mm



(e) Kante 0610-2, RGB-Bild.

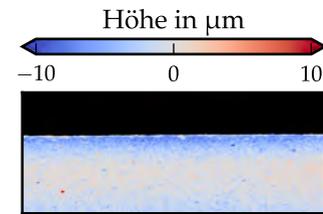


(f) Kante 0610-2, 3D-Punktwolke.

– 1 mm



(g) Kante 0610-4, RGB-Bild.



(h) Kante 0610-4, 3D-Punktwolke.

Abbildung 6.11 RGB-Bilder und 3D-Punktwolken ausgewählter Kanten. Bei 0407-2 werden die Grathöhe und -breite durch einen dicken Spritzer verfälscht. Der Grat von Kante 0446-2 hat einzelne Perlen, die ungleichmäßig verteilt sind. Die Kanten 2 und 4 von Teil 0610 weisen sehr unterschiedliche Gratausprägungen auf.

ten. Deshalb wurde untersucht, ob ein systematischer Zusammenhang zwischen der Grathöhe und einer der folgenden Einflussgrößen besteht:

- Schnittrichtung
- Walzrichtung
- Blechtafel
- Position auf dem Blech (Zeilen, Spalten, siehe Abbildung 4.2(b))
- Zeitpunkt des Schneidens, Maschinenlaufzeit
- Prozessparameter.

Eine eindeutige Einflussgröße konnte nicht identifiziert werden. Details der Analyse sind in Anhang A.5 zu finden.

6.1.4.2 Zusammenfassung

Aus der 3D-Punktwolke des Grats wurde die Grathöhe pro Kante in fünf Bereichen ermittelt. Während bei der Rautiefe Messfehler zu großen Streuungen der Rz -Werte führen, stellt die Messqualität bei der Gratummessung ein kleineres Problem dar. Es kommt dennoch – prozessbedingt – zu großen Unterschieden der Grathöhe entlang einer Kante, aber vor allem zwischen den vier Kanten desselben Teils. Die Ursache dafür konnte nicht eindeutig ermittelt werden. Deshalb wäre es denkbar, dass eine nicht erfasste Einflussgröße verantwortlich ist. Tritt der Laserstrahl nicht stabil im Zentrum der Düse aus dem Schneidkopf aus, sondern verändert seine Position, könnte dies beispielsweise die Grathöhe beeinflussen. Vermutlich würde eine größere Datenbasis in Kombination mit der Erfassung weiterer Einflussgrößen (z. B. Strahlzentrizität, exakte Position der Kanten bezüglich der Stegspitzen) helfen, die Ursache zu identifizieren.

Dass keine klaren Zusammenhänge zwischen dem Grat und den im vorigen Abschnitt genannten Einflussgrößen gefunden werden konnten, stellt für die Modellierung ein großes Problem dar. Für die eindeutige Zuordnung eines Teils (bzw. einer Parameterkombination) zu einer Grathöhe sollten die vier Kanten ähnliche Grathöhen aufweisen, was nicht der Fall ist. Deshalb sind die Ergebnisse eines Modells für eine einzelne Kante später nur begrenzt aussagefähig.

Da die Gratbreite stark mit der Grathöhe korreliert (Korrelationskoeffizient 0,7), lassen sich die Ergebnisse näherungsweise auf die Gratbreite übertragen. Darauf wird nicht weiter eingegangen.

6.2 Qualitätsbeurteilung

Eine objektive, überall anwendbare Qualitätsdefinition ist für die Formulierung individueller Anforderungen unabdingbar. Trotz existierender Normvorgaben wird die Schnittkantenqualität in der Praxis subjektiv von Einzelpersonen beurteilt. Im nächsten Abschnitt wird dieses Vorgehen kritisch hinterfragt und es wird untersucht, welche Merkmale der Schnittkante die subjektive Einschätzung des Menschen beeinflussen. Anschließend wird vorgestellt, wie die wichtigste Normgröße, die Rautiefe, mithilfe eines CNNs basierend auf einem Foto der Schnittkante geschätzt werden kann. Damit kann die menschliche Bewertung durch eine objektive Größe ergänzt werden, ohne dass ein teures Messgerät erforderlich ist.

6.2.1 Expertenbefragung

Fünf Experten (E1, E2, E3, E4 und E5) haben 100 Schnittkanten innerhalb von drei Wochen bewertet. In jeder Woche mussten sie die Qualität von 40 Kanten, jeweils separat für die Schnittfläche und den Grat, mit Schulnoten bewerten. Zehn der Kanten waren ohne das Wissen der Befragten in jeder Woche dieselben. In Abbildung 6.12 ist für jede der sechs Qualitätsstufen eine exemplarische Schnittfläche gezeigt. Zusätzlich zum Median sind die von den Einzelpersonen vergebenen Noten angegeben.

Die mit sehr gut bewertete Kante weist eine homogene, glatte Schnittfläche mit fast senkrechten Riefen auf. Die mit gut bewertete Kante unterscheidet sich davon nur durch kleinere Spritzer im mittleren Bereich. Die Riefen der mit Note 3 und 4 beurteilten Kanten verlaufen schräger, die Rauheit nimmt vor allem im unteren Bereich zu. Dadurch wirken die Kanten deutlich inhomogener. Die mit der Note 4 bewertete Kante ist zusätzlich an der Unterkante verfärbt. Ausgeprägte horizontale Schichten und dicke Spritzer sind auf den beiden schlecht bewerteten Kanten zu

— 1 mm



(a) Kante 0275-1, Note 1 (2, 1, 1, 1, 2).

— 1 mm



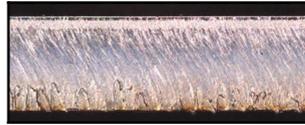
(b) Kante 0401-1, Note 2 (2, 1, 2, 2, 2).

— 1 mm



(c) Kante 0960-1, Note 3 (3, 3, 2, 2, 3).

— 1 mm



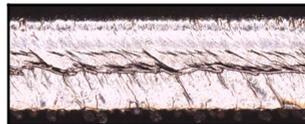
(d) Kante 0009-1, Note 4 (3, 4, 4, 4, 3).

— 1 mm



(e) Kante 0575-1, Note 5 (4, 5, 5, 6, 6).

— 1 mm

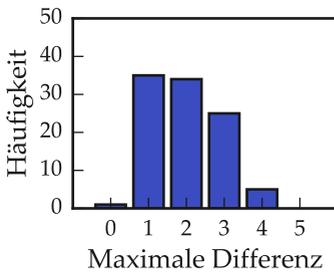


(f) Kante 0727-1, Note 6 (5, 6, 6, 6, 6).

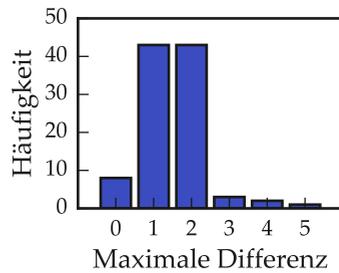
Abbildung 6.12 Beispiele für Schnittflächen verschiedener Qualitätsstufen. Zu jeder Kante sind der Median der fünf Bewertungen sowie die Noten der Einzelpersonen angegeben.

finden. Diese Merkmale sind bei der mit der Note 6 beurteilten Kante noch ausgeprägter als bei der Kante mit Note 5.

Um zu beurteilen, wie gut/schlecht eine Einzelperson für eine repräsentative Beurteilung der Qualität geeignet ist, ist in Abbildung 6.13 die Verteilung der maximalen Differenz, also der Differenz zwischen schlechtester und bester der fünf vergebenen Bewertungen derselben Kante, gezeigt. Bei lediglich einer Kante haben alle Experten dieselbe Note für die Schnittfläche vergeben. In den meisten Fällen kommt es zu Abweichungen von ein bis zwei Notenpunkten. Bei einem Drittel der Schnittflächen liegen die Bewertungen jedoch um drei bis vier Schulnoten auseinander. Für die Schnittfläche der Kante 0298-1 (siehe Abbildung 6.14(a)) vergaben die Experten die Noten 5, 5, 2, 2 und 6. Als Hauptgrund ihrer Bewertung nannten alle die Verfärbung/Oxidation des unteren Kantenbereichs. Die Schnittfläche der Kante 0857-1 (siehe Abbildung 6.14(b)) wurde mit 5, 4, 3,

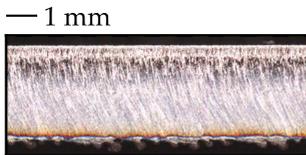


(a) Bewertung der Schnittfläche.

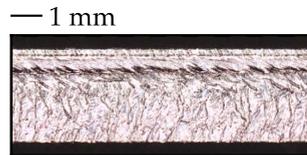


(b) Bewertung des Grats.

Abbildung 6.13 Verteilung der maximalen Differenz der fünf Expertenbewertungen derselben Kante.



(a) Kante 0298-1, Noten: 5, 5, 2, 2, 6.



(b) Kante 0857-1, Noten: 5, 4, 3, 6, 6.

Abbildung 6.14 Beispiele für Schnittkanten mit stark abweichenden Bewertungen der Schnittfläche.

6 und 6 bewertet. Die genannten Gründe beziehen sich alle auf die Rauheit der Kante. Beim Grat fallen die Beurteilungen einheitlicher aus. In 93 % der Fälle beträgt die Abweichung zwei oder weniger Notenpunkte.

Diskussion Differenzen von ein bis zwei Noten sind unproblematisch, da eine Differenzierung in sechs verschiedene Qualitätsklassen in der Praxis oft nicht erforderlich ist. Unterschiede von drei oder mehr Notenstufen bedeuten jedoch, dass die Qualitätseinschätzungen verschiedener Personen stark voneinander abweichen. An den beiden Beispielen mit sehr unterschiedlichen Bewertungen zeichnet sich bereits ab, was sich später bestätigen wird: Die Experten nehmen dieselben Merkmale wahr, aber sie gewichten diese unterschiedlich und kommen so zu einer abweichenden Einschätzung der Qualität. Die Beurteilungen des Grats fallen vermutlich einheitlicher aus, weil diese Struktur deutlich weniger komplex ist. Ein niedriger Grat wird positiv, ein hoher Grat negativ bewertet.

Tabelle 6.3 Maximale Notendifferenz bei der Bewertung der jeweils selben Kante durch dieselbe Person in drei verschiedenen Wochen. Für jede Kante wurde in jeder Woche außerdem der Median der fünf Bewertungen ermittelt. Dessen Schwankungen über die drei Wochen sind in der Tabelle ebenfalls angegeben.

	Differenz	E1	E2	E3	E4	E5	Median
Fläche	0	3	3	3	3	3	4
	1	3	4	4	6	3	6
	2	4	2	1	1	1	0
	3	0	1	2	0	2	0
	4	0	0	0	0	1	0
Grat	0	1	1	2	4	3	2
	1	6	4	7	5	0	6
	2	3	5	1	1	7	2

Die Bewertungen einer Schnittkante hängen nicht nur von der bewertenden Person, sondern auch von ihrer individuellen Tagesform ab. Dies lässt sich anhand der zehn wiederholt bewerteten Kanten feststellen, für die drei Bewertungen der Schnittfläche und des Grats von jeder Person vorliegen. In Tabelle 6.3 ist aufgelistet, wie oft sich die Bewertungen derselben Kante um null, eine, zwei oder drei Notenstufen unterscheiden. Bei zwei der fünf Personen kommt es bei der Beurteilung der Schnittfläche zu Abweichungen von drei, bei einer Person sogar von vier Notenstufen. Eine Abweichung von 3 bedeutet beispielsweise, dass von derselben Person in einer Woche die Note 2 und in einer anderen Woche für dieselbe Kante eine 5 vergeben wurde. So große Abweichungen treten bei der Gratsbewertung nicht auf.

Fazit Es zeigt sich, dass einzelne Personen aufgrund ihrer individuellen Präferenzen und Tagesform nur bedingt in der Lage sind, eine zuverlässige, reproduzierbare und repräsentative Einschätzung der Schnittflächenqualität vorzunehmen. Entweder ermöglicht man es also, individuelle Qualitätspräferenzen anzugeben und passt die Prozessparameter entsprechend an (mehr dazu in Abschnitt 6.4.5), oder man verwendet einen Durchschnittswert, der die Einschätzung mehrerer Personen wiedergibt.

Dafür wird nun der Median der zu jeder Kante vorliegenden Bewertungen genutzt. Tabelle 6.3 zeigt, dass der Median relativ stabil ist. Bei der Schnittfläche schwankt er zwischen den drei Wochen um maximal einen Notenpunkt, beim Grat in den meisten Fällen ebenfalls. Nur bei zwei von zehn Kanten kommt es zu einer Abweichung um zwei Notenpunkte. Vergleicht man die Bewertungen der einzelnen Experten für alle 100 bewerteten Schnittflächen mit dem jeweiligen Median, zeigt sich, dass sie verschieden kritisch sind. Die Experten 1 und 5 vergeben durchschnittlich schlechtere Noten als die anderen. E1 vergab 33-mal eine schlechtere und nur 12-mal eine bessere Note als der Median. E5 bewertete 32 Kanten schlechter und nur 10 Kanten besser. Bei E2 halten sich die besseren (26) und die schlechteren (20) Bewertungen die Waage. Die Experten 3 und 4 sind weniger kritisch. E3 bewertete 31 Kanten besser und 15 schlechter als der Median, E4 beurteilte 49 Kanten besser und 14 Kanten schlechter. Bei der Bewertung des Grats ist die Tendenz dieselbe.

Zusätzlich zu den Noten liegen zu 70 Kanten die Einschätzungen jedes Experten für fünf Merkmale der Schnittfläche vor. Diese sind: Rauheit, Rechtwinkligkeit, Verfärbung, Homogenität und Anschmelzung der Oberkante. Die Experten bewerteten jeweils, ob ihnen dieses Merkmal gut, mittelmäßig oder schlecht gefiel. Um einen Einblick in die individuelle Gewichtung dieser fünf Merkmale zu erhalten, soll für jeden Experten quantifiziert werden, wie sich seine Wahrnehmung der Merkmale auf die Note auswirkt. Dafür sind in Tabelle 6.4 die Korrelationen zwischen der Note der Schnittfläche und der Beurteilung der Einzelkriterien dargestellt. Besonders interessant ist der Vergleich der Experten E1 und E4. Während für E1 alle Kriterien bis auf die Verfärbung relevant sind, sind für E4 lediglich die Rauheit und die Homogenität von Bedeutung. Vermutlich kommt E1 deshalb oft zu schlechteren, E4 oft zu besseren Bewertungen als der Durchschnitt. Die Rauheit und die Homogenität sind für alle fünf Personen wichtig, die Rechtwinkligkeit scheint nur für E1 eine Rolle zu spielen. Die Verfärbung gewichten die Experten E2 und E5 stärker als die anderen, bei der Beschaffenheit der Oberkante trifft das auf E1 und E3 zu. Durch die individuelle Gewichtung der einzelnen Merkmale kommen die fünf Personen zu sehr unterschiedlichen Ergebnissen in der Gesamtbewertung. Kante 0298-1 (siehe Abbildung 6.14(a)) wird von allen Experten als verfärbt wahrgenommen. Da dieses Merkmal für E3 und E4 irrelevant ist, geben

Tabelle 6.4 Pearson-Korrelationskoeffizienten zwischen Schulnotenbewertung der Schnittfläche und der Beurteilung von fünf Einzelmerkmalen: Rauheit, Rechtwinkligkeit, Verfärbung, Homogenität und Anschmelzung der Oberkante. Ergänzend sind die Korrelationskoeffizienten mit der gemessenen, durchschnittlichen Rautiefe und der Rechtwinkligkeitstoleranz (u) dargestellt. In der letzten Zeile stehen die Korrelationskoeffizienten zwischen der Schulnotenbewertung des Grats und der gemessenen Grathöhe.

Einzelkriterien	E1	E2	E3	E4	E5
Rauheit	0,6	0,7	0,7	0,8	0,6
Rechtwinkligkeit	0,5	0,3	–	0,0	0,2
Verfärbung	0,2	0,4	0,0	0,1	0,4
Homogenität	0,6	0,7	0,7	0,8	0,6
Oberkante	0,5	–	0,4	0,1	0,3
Rautiefe	0,5	0,5	0,7	0,8	0,6
u	0,3	0,3	0,6	0,6	0,4
Grathöhe	0,8	0,7	0,9	0,8	0,8

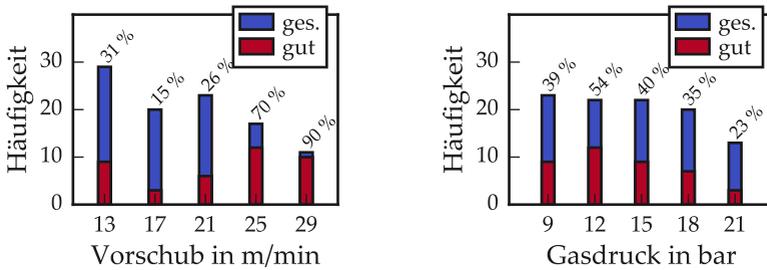
sie die Note 2, während die anderen die Note 5 oder 6 vergeben. Ebenfalls in der Tabelle 6.4 abzulesen sind die Korrelationskoeffizienten zwischen der Note und der gemessenen (über alle Messwerte der Kante gemittelten) Rautiefe bzw. der gemessenen Rechtwinkligkeitstoleranz. Die Korrelation zwischen Note und wahrgenommener Rauheit ist in zwei Fällen größer als die zwischen Note und gemessener Rautiefe, in drei Fällen gleich. Abweichungen könnten daran liegen, dass es sich bei der gemessenen Rautiefe hier lediglich um einen Mittelwert handelt, der Unterschiede innerhalb einer Kante unzureichend wiedergibt. Dass die Korrelation zwischen der Note und der gemessenen Rechtwinkligkeitstoleranz ausgerechnet bei den Experten am größten ist, denen die wahrgenommene Rechtwinkligkeit am wenigsten wichtig ist (E3, E4), kann anhand dieser Daten nicht erklärt werden. In der letzten Zeile der Tabelle 6.4 stehen die Korrelationskoeffizienten zwischen Gratnote und gemessener Grathöhe. Hier gibt es einen sehr eindeutigen Zusammenhang: Die Korrelationskoeffizienten liegen zwischen 0,7 und 0,9.

Zuletzt wird untersucht, wie die Qualität mit den Kosten eines Teils zusammenhängt. Da die Kosten hauptsächlich vom Vorschub (Zeit) und vom Gasdruck (Gasverbrauch) abhängen, werden in Abbildung 6.15 die

guten Kanten hinsichtlich dieser Parameter mit allen bewerteten Kanten verglichen. Wie man sieht, wurden die meisten guten Kanten mit einem gegenüber der LTT erhöhten Vorschub und verringerten Gasdruck geschnitten. Die einzige Kante, welche sowohl für die Schnittfläche als auch für den Grat die Bestnote erhalten hat, ist die Kante 0275-1 (siehe Abbildung 6.12(a)). Sie wurde mit einem Vorschub von 25 m/min und einem Gasdruck von 18 bar geschnitten. Die Experten bewerteten auch die Kante 0401-1, die mit LTT-Parametern geschnitten wurde. Sie erhielt die Noten 2 (Schnittfläche) und 3 (Grat).

Diskussion Durch eine Erhöhung des Vorschubs gegenüber dem LTT-Wert könnten sowohl die Produktivität gesteigert als auch die Qualität verbessert werden. Bei der Wahl der LTT-Parameter geht es jedoch nicht nur um die Maximierung der Qualität, sondern auch darum ein besonders stabiles Prozessfenster zu finden, damit es in der industriellen Anwendung keinesfalls zu Fehlschnitten kommt. Dafür nimmt man bewusst eine Verringerung der Produktivität und eine Verminderung der Qualität in Kauf. Bislang werden Produktivitätssteigerungen vor allem durch immer höhere Laserleistungen erzielt, was mit einem großen Entwicklungsaufwand einhergeht. Eine Verbesserung der Kontrollierbarkeit des Gesamtprozesses würde hier neue Möglichkeiten eröffnen, die Produktivität zu erhöhen.

In der Tabelle 2.1 sind die Qualitätsklassen thermisch geschnittener Kanten anhand der maximalen Rautiefe und Rechtwinkligkeitstoleranz definiert. Eine Kante der Dicke 3 mm ist bei einem Rautiefenwert unterhalb von 11,8 μm der Klasse 1 zuzuordnen. Für Qualitätsklasse 2 darf die Rautiefe 42 μm nicht überschreiten. Keine der im Rahmen der Expertenbefragung insgesamt mit gut bewerteten Kanten erfüllt die Kriterien von Klasse 1 hinsichtlich der Rautiefe, was beim Laserschneiden eines Edelstahlblechs dieser Dicke mit einem Festkörperlaser auch zu erwarten ist. Die Werte liegen zwischen 15 μm (Kante 0631-1) und 38 μm (Kante 0314-1) und damit in Klasse 2. Für eine Zuordnung zu Klasse 1 muss die Rechtwinkligkeitstoleranz kleiner als 59 μm sein. Diese Bedingung erfüllen alle gut beurteilten Kanten mit Werten zwischen 7 μm (Kante 1040-1) und 58 μm (Kante 1013-1).



(a) Häufigkeit der Vorschubstufen jeweils für alle Kanten (ges.) und die mit gut bewerteten Kanten. Der LIT-Wert ist 21 m/min. (b) Häufigkeit der Gasdruckstufen jeweils für alle Kanten (ges.) und die mit gut bewerteten Kanten. Der LIT-Wert beträgt 15 bar.

Abbildung 6.15 Zusammenhang zwischen der Note und den Parametern Vorschub und Gasdruck. Bei den guten Kanten wurden sowohl die Schnittfläche als auch der Grat mit einem Median von 2 oder besser bewertet.

6.2.1.1 Fazit

Aus der Auswertung der Expertenbefragung lassen sich die folgenden Schlüsse ziehen:

- Verschiedene Menschen haben unterschiedliche Präferenzen. Das gilt vor allem für die Schnittfläche, die verschiedene Merkmale aufweist, welche unterschiedlich gewichtet werden. Beim Grat weichen die Bewertungen weniger voneinander ab. Die Qualitätsbewertung sollte nicht einer einzelnen Person überlassen werden, da diese weder ein repräsentatives noch ein reproduzierbares Urteil fällen kann.
- Wichtig für die Schnittflächenbewertung aller Experten sind die Rauheit und die Homogenität. Ob eine Schnittkante homogen aussieht, hängt vermutlich davon ab, wie stark sich die verschiedenen horizontalen Schichten unterscheiden und wie deutlich einzelne Riefen sichtbar sind. Es besteht also ein Zusammenhang mit der in verschiedenen Blechtiefen gemessenen Rautiefe. Die Rechtwinkligkeitstoleranz scheint in dem 3 mm dicken Edelstahlblech eine sehr untergeordnete Rolle für das menschliche Qualitätsurteil zu spielen, deshalb wird sie im weiteren Verlauf der Arbeit nicht mehr thematisiert. In anderen Blechdicken kann das anders sein.

- Eine Ergänzung der subjektiven Einschätzung durch objektive Kriterien, insbesondere Rautiefe und Grathöhe, wäre sehr sinnvoll. Für eine Nachbildung des (durchschnittlichen) menschlichen Eindrucks reichen diese beiden Größen jedoch nicht aus. Im Rahmen der CIRP-Konferenz Intelligent Computation in Manufacturing Engineering 2020 wurde ein Modell präsentiert, mit dem der Median der Expertenbewertungen basierend auf objektiven Merkmalen der Schnittkante nachgebildet werden kann [TP20c]. Als Eingangsdaten dienten neben den bereits eingeführten Messgrößen zahlreiche weitere Merkmale, die aus den 3D-Punktwolken berechnet wurden. Der subjektive Eindruck ließ sich zuverlässig nachbilden. Die Interpretation der Zusammenhänge zwischen Merkmalen und menschlichem Eindruck erwies sich trotz eines einfachen Modellansatzes aufgrund der Merkmalskomplexität als schwierig. Für eine allgemeingültige Qualitätsbewertung reicht die aktuelle Datenbasis noch nicht aus, weil lediglich fünf Personen desselben Unternehmens befragt wurden.
- Geschnittene Teile werden in der Praxis oft nachbearbeitet, um die Qualität zu verbessern (siehe Abschnitt 2.2.2). Je nachdem welche Nachbearbeitungsmethoden vorhanden sind, könnten die Präferenzen hinsichtlich der Schnittkante noch unterschiedlicher sein. Muss jedes Teil z. B. manuell entgratet werden, sollte die Minimierung des Grats angestrebt werden. Ist dagegen eine Entgratmaschine vorhanden, könnte die maximale Produktivität das wichtigste Kriterium sein.
- Wegen der individuellen Präferenzen, der unterschiedlichen Möglichkeiten zur Nachbearbeitung und der Vielzahl von Einsatzbereichen für mit dem Laser geschnittene Teile ist der Mehrwert eines Modells, das die Qualitätsbewertung eines „Durchschnittsmenschen“ nachbilden kann, vermutlich eher gering. Nützlicher wäre die Möglichkeit, den Schneidprozess automatisiert an individuelle Kundenanforderungen anpassen zu können.
- Bei einem gegenüber dem LTT-Wert erhöhten Vorschub sind besonders viele gut bewertete Schnittkanten zu finden. Dieser Bereich wird bislang aus Sicherheitsgründen nicht genutzt. Wenn man den

Tabelle 6.5 Einschätzung der Experten, welcher Prozessparameter kleiner, gleich oder größer als der LTT-Wert eingestellt war. Die Anzahl der richtig klassifizierten Parameter (für insgesamt 10 Teile) ist hier angegeben.

Experte	ADB	EM	Gasdruck	Vorschub
E1	5	6	5	3
E2	–	–	–	–
E3	3	2	3	1
E4	2	4	0	4
E5	–	–	–	–
∅	33 %	40 %	26 %	26 %

Schneidprozess besser kontrollieren könnte und kleinere Sicherheitsfaktoren vorhalten müsste, würde man eine deutliche Produktivitätssteigerung bei gleichzeitiger Verbesserung der Qualität erreichen. Die in Abschnitt 6.3 beschriebene Methode zur Überwachung des Maschinenzustands trägt zu einer besseren Kontrollierbarkeit bei.

6.2.1.2 Schätzung der Prozessparameter durch die Experten

In einer Woche der Expertenbefragung waren die Experten aufgefordert, für zehn Teile einzuschätzen, wie die zugrunde liegenden Prozessparameter eingestellt waren. Dabei gab es für jeden Parameter drei Auswahlmöglichkeiten: kleiner als der LTT-Wert, gleich dem LTT-Wert und größer als der LTT-Wert. Die Ergebnisse sind in Tabelle 6.5 zusammengefasst.

Zwei der fünf Experten gaben an, keine Aussage treffen zu können. Die anderen drei klassifizierten die zehn Teile. Beim ADB lagen sie durchschnittlich bei 33 % der Fälle richtig, beim EM bei 40 % der Teile und beim Gasdruck und Vorschub ordneten sie 26 % der richtigen Klasse zu. Eine rein zufällige Wahl der Klasse führt bei drei Auswahlmöglichkeiten zu einer Trefferquote von 33 %, was ungefähr der Leistung der Experten entspricht.

Diskussion Auch wenn die Anzahl der befragten Personen und die Anzahl zu klassifizierender Teile zu klein war, um zu fundierten Aussagen zu kommen, kann man davon ausgehen, dass Menschen nicht in der Lage

sind, von der Schnittkante zurück auf die Prozessparameter zu schließen. Einige der Experten glaubten, dass das prinzipiell nicht möglich sei, weil es keinen eindeutigen Zusammenhang zwischen Kante und Prozessparametern gebe und ganz unterschiedliche Parameterkombinationen zum selben Schnittbild führen würden. Wie in Abschnitt 6.3.1 gezeigt wird, ist diese Annahme falsch. Die Komplexität der Zusammenhänge war lediglich zu groß für die befragten Experten und ist es vermutlich auch für alle anderen Menschen.

6.2.2 Bildbasierte Schätzung der Rautiefe

Wie an anderen Stellen bereits erläutert (siehe z. B. Abschnitt 1.1.1 und 6.2.1.1), erscheint eine bildbasierte Schätzung der Rautiefe als sinnvoll, damit diese Größe ohne Oberflächenmessgerät zur Verfügung gestellt werden kann. Deshalb wird hier ein Ansatz zur Bestimmung des Rz -Wertes an verschiedenen Stellen der Schnittkante mit einem CNN vorgestellt. Ähnliche Ergebnisse wurden bereits im Rahmen der CIRP-Konferenz Photonic Technologies 2020 vorgestellt [TP20a].

6.2.2.1 Einfluss des Datentyps

Es wurde bereits darauf hingewiesen, dass eine schlechte Datenqualität negative Auswirkungen auf datenbasierte Modelle haben kann. Auf diesen Aspekt wird folglich ein besonderes Augenmerk gelegt. Für die bildbasierte Bestimmung der Rautiefe wurde das in Abschnitt 5.1.3 beschriebene CNN verwendet. Als Eingangsdaten wurden die vorverarbeiteten quadratischen RGB-Bilder der Schnittkante genutzt (siehe Abschnitt 4.2.2.1). Als Labels wurden für jedes Bild neun Rz -Werte (der Messlinien 03 bis 27) verwendet. Für Training und Validierung wurden einmal die mit dem Autoencoder korrigierten (CNN AE) und einmal die originalen, unbereinigten Messdaten (CNN MI) eingesetzt. Getestet wurde in beiden Fällen mit den Rz -Werten, die sich aus der Kombination der Messdurchgänge I und II ergeben, weil dies die qualitativ besten verfügbaren Daten sind (siehe Abbildung 6.5). Diese wurden nicht für das Training eingesetzt, weil hier die Beurteilung der praktischen Anwendbarkeit im Vordergrund steht. Dafür sollte jede Schnittkante aus Zeitgründen nur einmal vermessen wer-

den müssen. Außerdem kann man so gut den Einfluss der Datenqualität auf die Modellgüte verdeutlichen.

In Abbildung 6.16 sind die mittleren absoluten Testfehler pro Messlinie gezeigt. Zusätzlich zu den MAEs der beiden CNNs sind die sogenannten theoretischen Fehler aufgetragen: Ein Modell hätte diese Fehler, wenn es während des Trainings lediglich den Mittelwert jeder Messlinie lernen und diesen beim Testen vorhersagen würde. Da das CNN AE und das CNN MI auf unterschiedlichen Datensätzen (gleiche Bilder, andere Labels) trainiert wurden, unterscheiden sich auch ihre theoretischen Fehler. Bei CNN AE liegt der MAE stets unterhalb des theoretischen Fehlers. Dieses Netz ist also in der Lage, die Rautiefenwerte ausgehend von einem RGB-Bild zu schätzen. Der über alle Messlinien gemittelte Fehler beträgt $2,7 \mu\text{m}$, den größten Fehler weist die Messlinie 06 mit $3,4 \mu\text{m}$ auf, den kleinsten die Messlinie 27 mit $2,3 \mu\text{m}$. Der Fehler im oberen Bereich der Schnittkante ist etwas größer als im mittleren und unteren Bereich. Bei CNN MI ist der theoretische Fehler größer, da der Mittelwert der unbereinigten Daten stärker von den zum Testen verwendeten Labels abweicht. Der Fehler von CNN MI liegt zwar ebenfalls unterhalb des theoretischen Fehlers, ist jedoch deutlich größer als der von CNN AE. Die Unterschiede der MAEs der beiden Netze sind für die Messlinien 03 und 06 besonders groß, der Fehler des auf den Originaldaten trainierten Modells ist aber für alle Messlinien höher. Der gemittelte Fehler beträgt $4,8 \mu\text{m}$. Eine bessere Trainingsdatenbasis führt also zu einer besseren Modellgüte.

Warum der Fehler von Messlinie 03 bei beiden Modellen besonders groß ist, wird in Abbildung 6.17 deutlich. Dort werden die Vorhersagen der beiden CNNs jeweils mit den Labels der Testdaten verglichen. Idealerweise sollten alle Datenpunkte auf der schwarzen Linie liegen. Das ist vor allem bei Messlinie 03 nicht der Fall. Dort kommt es bei CNN AE zu größeren Abweichungen zwischen Vorhersage und Label aufgrund von zu kleinen vorhergesagten Rauheitswerten. Bei CNN MI kommt es ebenfalls zu größeren Abweichungen, allerdings sind die vorhergesagten Werte in den meisten Fällen deutlich größer als das Label. Für die Messlinien 15 und 27 erreichen beide Modelle eine ähnliche Güte. Es treten nur wenige große Abweichungen auf, bei CNN MI etwas mehr als bei CNN AE.

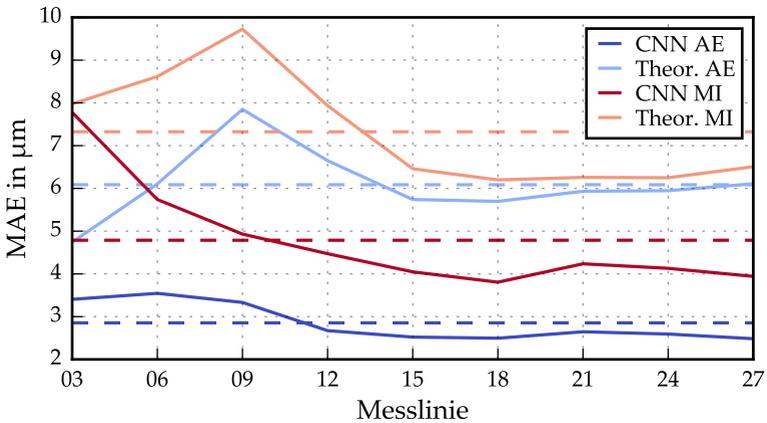


Abbildung 6.16 Mittlerer absoluter Fehler pro Messlinie auf dem Testdatensatz von Split 1. Es sind die Fehler des CNNs AE und des CNNs MI aufgetragen, die mit verschiedenen Labels trainiert, aber mit denselben Daten getestet wurden. Zusätzlich ist der theoretische Fehler eingetragen. Die gestrichelten Linien zeigen den Mittelwert über die neun Messlinien der jeweiligen Kurve an.

Diskussion Die Vorhersagen von CNN AE sind nur auf den ersten Blick teilweise falsch und viel zu niedrig: Beispielsweise werden für Messlinie 03 bei Labels über $40\ \mu\text{m}$ Rautiefenwerte unter $20\ \mu\text{m}$ vorhergesagt. Alle in Abbildung 6.17(a) gezeigten Datenpunkte, die in der Kombination der Messdurchgänge I und II einen Wert von über $40\ \mu\text{m}$ aufweisen, wurden manuell überprüft. Sie sind (fett gedruckt), zusammen mit den Rautiefenwerten der anderen drei Kanten desselben Teils, in Tabelle 6.6 aufgelistet. Wie man sieht, weisen mindestens zwei der Kanten desselben Teils deutlich kleinere Rautiefenwerte auf. Deshalb ist es gut möglich, dass die Labels in diesem Fall fälschlicherweise etwas zu groß sind und dass das CNN AE den Rautiefenwert richtig einschätzt. Die großen Abweichungen würden in diesen Fällen also nicht durch falsch gelernte Modellparameter, sondern durch die (immer noch) verbesserungswürdige Datenqualität verursacht.

Wie bereits in den Abschnitten 4.2.1.3 und 6.1.2 beschrieben, kommt es bei der Rz -Messung aufgrund der hohen Reflektivität der Edelstahloberflächen häufig zu Problemen und zu falschen, viel zu großen Messwerten.

Tabelle 6.6 Rautiefenwerte ausgewählter Kanten in der Kombination der Messdaten von Messung I und II. Die zugehörigen Bilder der Schnittkanten sind in Anhang A.3 zu finden.

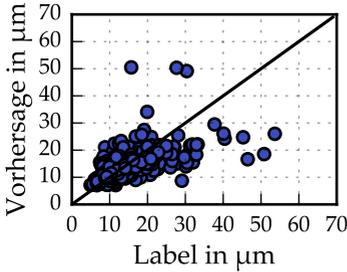
Teil	Kante 1 in μm	Kante 2 in μm	Kante 3 in μm	Kante 4 in μm
0949	9	11	54	38
0992	12	10	51	16
1006	12	9	40	33
1056	8	8	13	47
1077	8	7	45	40

Bei der Kombination der beiden Messdurchgänge werden viele, aber nicht alle Messfehler (ausreichend) korrigiert. Das CNN AE konnte, obwohl seine Trainingsdatenbasis nicht fehlerfrei war, den grundlegenden Zusammenhang zwischen dem Bild einer Schnittkante und dem Rautiefenwert an verschiedenen Stellen erlernen. Die Modellgüte wird möglicherweise aufgrund von falschen Testlabels sogar unterschätzt. Bei CNN MI ist das anders. Hier treten tatsächlich falsche, zu große Vorhersagen aufgrund nicht korrekt erlernter Zusammenhänge zwischen Bild und Rautiefe auf. Der Grund dafür ist die schlechtere Trainingsdatenbasis. Die unbereinigten Messdaten des ersten Messdurchgangs enthalten mehr und stärkere Ausreißer als die mit dem Autoencoder korrigierten Messwerte. CNN AE konnte die systematischen Zusammenhänge erlernen, bei CNN MI war das kaum noch möglich. Die Vorhersagen beider CNNs sind besonders mangelhaft für die Messlinien 03 und 06, deren Datenqualität am schlechtesten ist.

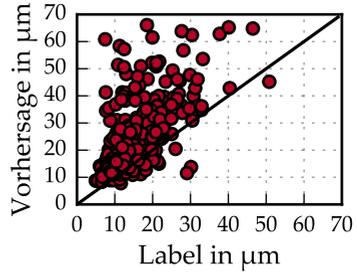
6.2.2.2 Einfluss der Datenaufteilung und Überanpassung

In Abbildung 6.18(a) sind die Fehler von drei CNNs gegenübergestellt. Sie wurden auf unterschiedlichen Teilmengen (Split 1, 2 oder 3) der mit dem Autoencoder bereinigten Daten trainiert und validiert und auf ebenfalls verschiedenen Teilmengen der kombinierten Messdaten von Durchgang I und II getestet. So lässt sich der Einfluss der Datenaufteilung untersuchen.

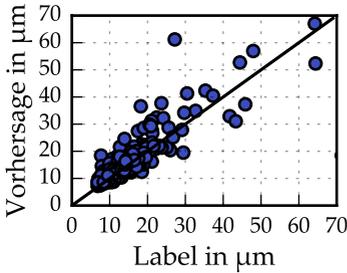
Die Modellfehler auf Testset 1 und 3 sind sehr ähnlich. Die größte Abweichung zwischen ihnen beträgt $0,7 \mu\text{m}$ bei Messlinie 09. Die MAEs auf



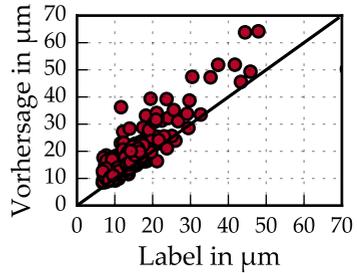
(a) Vorhersage von CNN AE und Label für Messlinie 03.



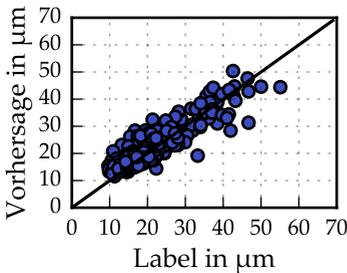
(b) Vorhersage von CNN MI und Label für Messlinie 03.



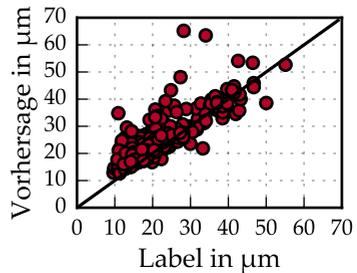
(c) Vorhersage von CNN AE und Label für Messlinie 15.



(d) Vorhersage von CNN MI und Label für Messlinie 15.

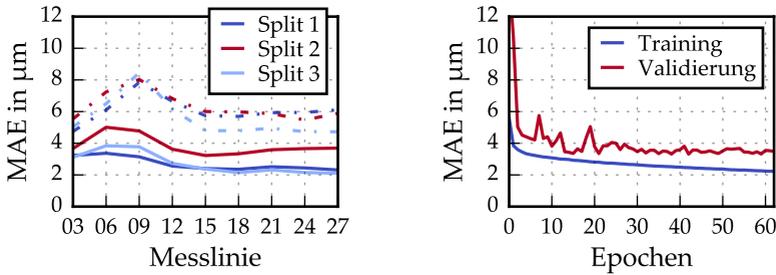


(e) Vorhersage von CNN AE und Label für Messlinie 27.



(f) Vorhersage von CNN MI und Label für Messlinie 27.

Abbildung 6.17 Vergleich von Vorhersage und Label auf dem Testdatensatz (Split 1) für die R_z -Messlinien 03, 15 und 27.



(a) Mittlerer absoluter Fehler pro Messlinie für drei CNNs AE, die auf verschiedenen Aufteilungen der Daten trainiert und getestet wurden. Gestrichelt eingetragen sind die theoretischen Fehler. (b) Trainings- und Validierungsfehler des CNNs AE auf Split 1.

Abbildung 6.18 Mittlerer absoluter Fehler für drei verschiedene Aufteilungen der Daten und Verlauf der Trainings- und Validierungsfehler über der Epochenanzahl.

dem Testset 2 sind systematisch größer, zwischen $0,5 \mu\text{m}$ auf Linie 03 bis zu $1,7 \mu\text{m}$ auf Messlinie 09. Die Trainings- und Validierungsfehler auf den drei Datensätzen entwickeln sich sehr ähnlich, nur die Testfehler unterscheiden sich.

Diskussion Warum der Testfehler bei Split 2 höher ausfällt, ist schwer zu sagen. Eine mögliche Erklärung ist, dass die Trainings- und Validierungsdaten aufgrund der kleinen Datenbasis bei manchen Aufteilungen nicht ausreichend repräsentativ sind, um die Zusammenhänge auf unbekanntem Daten korrekt vorherzusagen.

In Abbildung 6.18(b) ist der Verlauf der Trainings- und Validierungsfehler – gemittelt über die neun Messlinien – für Split 1 gezeigt. Der Validierungsfehler schwankt zunächst, stabilisiert sich dann aber. Mit zunehmender Epochenanzahl kommt es zu einer Überanpassung an die Trainingsdaten. Der Validierungsfehler beträgt $3,3 \mu\text{m}$, der Trainingsfehler $2,5 \mu\text{m}$.

6.2.2.3 Fazit

Auch wenn es auf den ersten Blick nicht so aussehen mag, weil es möglich war, die grundsätzlichen Zusammenhänge zwischen RGB-Bild und Rauftiefe zu modellieren, sind die Messfehler in vielerlei Hinsicht ein Problem.

Sie vergrößern den MAE sowohl während des Trainings als auch im Test. Aufgrund ihrer Größe können sie dazu führen, dass das Modell während des Trainings versucht, die Ausreißer „auswendig“ zu lernen, anstatt die generellen Zusammenhänge zu verstehen, weil sich die Verlustfunktion dadurch deutlich verringern lässt. Dadurch wird das Netz an die Trainingsdaten überangepasst. Je nachdem wie viele Messfehler in den Daten auftreten, können die grundlegenden Zusammenhänge gut (CNN AE) oder eher schlecht (CNN MI) erlernt werden.

Auch wenn hauptsächlich die oberen Messlinien fehlerbehaftet sind, betrifft das Problem die Vorhersage auf allen Messlinien, weil die Fehler der neun Regressionsziele während des Trainings gemittelt werden. Das könnte der Grund dafür sein, dass die Fehler des CNNs MI grundsätzlich größer sind als die von CNN AE. Der wichtigste Schritt zur Verbesserung des Modells ist deshalb die weitere Verbesserung der Datenqualität. Der zuverlässigste Weg wäre es, Schnittkanten mehrfach aufzunehmen. Um dies effizienter zu gestalten, könnten beispielsweise zunächst nur die Kanten erneut vermessen werden, bei denen die Abweichung zwischen Label und Modellvorhersage besonders groß ist. Noch besser wäre es, die R_z -Labels mit einem tastenden Messverfahren zu generieren. Zusätzlich könnte durch eine stärkere Regularisierung der Modellparameter eine geringere Anpassung an einzelne, fehlerhafte Trainingsbeispiele erreicht werden.

6.3 Überprüfung des Maschinenzustands

Im Folgenden wird zunächst das Konzept zur Überprüfung des Maschinenzustands vorgestellt. Dieses basiert auf der Schätzung der Ist-Prozessparameter mit einem faltenden neuronalen Netz. Die Güte des Modells wird im Anschluss detailliert evaluiert.

In Abschnitt 2.4.2 wurde erläutert, wie der Zustand einer Laserschneidmaschine in der Praxis überprüft wird. Ein klar definierter und stabiler Maschinenzustand ist die Grundvoraussetzung für die Beherrschbarkeit des Prozesses, weil Laserschneiden ein gesteuerter Prozess ist, bei dem Soll- und Ist-Parameter (ausreichend gut) übereinstimmen müssen. Es gibt zwar erste Ansätze zur Regelung des Schneidprozesses, beispielsweise von Hesse et al. [HSR14] und Thombansen et al. [Tho+14], aber

diese erfordern zusätzliche Hard- und Software und sind bislang in fast keiner Maschine vorhanden. Deshalb wurde ein Konzept entwickelt, mit dem der Zustand einer Laserschneidmaschine automatisiert überprüft bzw. dauerhaft überwacht werden kann. Anstatt Sensorsignale von Einzelkomponenten auszuwerten, wird dafür die Schnittkante verwendet, weil dort alle Einflussgrößen zusammenkommen. Ausgehend von einem Bild der Schnittkante werden die Ist-Prozessparameter mit einem CNN bestimmt und mit den eingestellten Soll-Werten verglichen. Anders als die genannten Regelungsansätze erfordert dieses Konzept keine aufwendige Sensorik, sondern lediglich ein RGB-Bild der Schnittkante. Eine Regelung kann damit jedoch nicht umgesetzt werden, da die Kante während des Schneidens nicht aufgenommen werden kann.

Solange sich die Maschine im Soll-Zustand befindet, stimmen beide Größen überein. Differenzen deuten auf Probleme in der Maschine hin. Wurde beispielsweise eine Fokusposition von $-1,0$ mm eingestellt, das neuronale Netz erkennt jedoch, dass sich der Strahlfokus an der Blechoberkante (bei $0,0$ mm) befindet, so verschiebt eine Maschinenkomponente den Strahlfokus nach oben, z. B. weil sie verschmutzt ist und sich erhitzt. Auch wenn es nicht möglich ist, von einer Parameterabweichung direkt auf die defekte oder verschmutzte Komponente zu schließen, schränkt die Kenntnis des veränderten Parameters den Suchraum erheblich ein und verkürzt die Stillstandszeit. In voll automatisierten Maschinen, in denen regelmäßig ein Bild der Schnittkante aufgenommen werden kann, könnte Ausschuss vermieden werden, indem die Maschine angehalten wird, sobald sich das Prozessergebnis verändert.

Dieser Ansatz funktioniert nur, wenn es möglich ist, die Ist-Parameter aus der Schnittkante abzuleiten. Dass der Mensch dazu nicht in der Lage ist, wurde in Abschnitt 6.2.1.2 dargestellt. Mit dem entwickelten CNN kann diese Aufgabe aber bewältigt werden. Eine Kurzfassung der präsentierten Ergebnisse wurde bereits publiziert [Tat+21].

6.3.1 Bildbasierte Regression der Prozessparameter

Für die Regression der fünf Prozessparameter ADB, EM, Fokusposition, Gasdruck und Vorschub wurde das in 5.1.1 beschriebene CNN entwickelt und trainiert. In Tabelle 6.7 ist der mittlere absolute Fehler des Modells für drei verschiedene Eingangsdattentypen gezeigt. Die RGB-, Höhen-

und RGBD-Bilder wurden jeweils dreimal (Split 1, 2 und 3) in Trainings-, Validierungs- und Testdatensatz aufgeteilt. Um die Größe der Fehler besser einschätzen zu können, ist der sogenannte theoretische Fehler ebenfalls angegeben. Dabei handelt es sich um den Fehler eines Modells, das im Training lediglich den Mittelwert des jeweiligen Parameters lernt und diesen beim Testen vorhersagt.

6.3.1.1 Einfluss des Datentyps

Die Ergebnisse zeigen, dass es möglich ist, die Prozessparameter direkt von der Schnittkante abzuleiten. Das CNN, welches auf den RGB-Bildern trainiert wurde, macht bei der Schätzung des Vorschubs einen Fehler von durchschnittlich 1,1 m/min. Das entspricht 23 % des theoretischen Fehlers. Für die Fokusposition beträgt der Fehler 0,2 mm. Das sind 22 % des theoretischen Fehlers. Die durchschnittlichen MAEs für ADB, EM und Gasdruck sind größer. Sie entsprechen 57, 44 und 51 % des theoretischen Fehlers. Die Fehler von ADB, EM und Fokusposition unterscheiden sich zwischen den drei Splits um maximal 0,1 mm, beim Gasdruck um 0,1 bar. Lediglich beim Vorschub kommt es zu einer etwas größeren Abweichung zwischen Split 1 bzw. 3 und 2.

Die über die Splits gemittelten Fehler für den ADB, das EM und die Fokusposition unterscheiden sich kaum zwischen den drei Eingangsdattentypen. Der Vorschub wird bei Vorhandensein der Höheninformation etwas besser geschätzt. Der durchschnittliche MAE ohne die Höheninformation beträgt 1,1 m/min und mit der Höheninformation 0,9 m/min. Lediglich beim Gasdruck liegt ein deutlicher Unterschied zwischen RGB- bzw. RGBD- und Höhen-Graustufenbild als Eingangsdaten vor. Wenn keine Farbinformation vorhanden ist, ist der Gasdruck schlechter einschätzbar. Mit der Farbinformation ist der durchschnittliche MAE 1,9 bar, ohne diese ist der Fehler um 0,4 bar größer.

Diskussion Vermutlich sind Vorschub und Fokusposition die dominanten Einflussgrößen und werden von den anderen Einflüssen am wenigsten überlagert. Deshalb sind sie für das CNN besonders gut zu erkennen. Schwieriger ist die Erkennung der beiden Parameter, die die Gasdynamik beeinflussen: ADB und Gasdruck. Da ADB, EM und Fokusposition zusammenhängen, pflanzt sich der Fehler bei der Bestimmung des ADBs

vermutlich auf das EM fort. Anders ist nicht zu erklären, dass die Fokusposition deutlich besser geschätzt werden kann als das Einstellmaß. Die Schwankungen der Fehler für verschiedene Aufteilungen der Daten sind relativ klein. Sie werden vermutlich durch die kleine Datenbasis verursacht. Wahrscheinlich beeinflusst der Gasdruck die Farbe der Schnittkante stärker als die anderen Parameter und kann deshalb bei Fehlen der Farbinformation etwas schlechter geschätzt werden. Diese Vermutung wird in Abschnitt 6.4.1 nochmals aufgegriffen. Dass die Höheninformation nicht zwingend benötigt wird, um die Prozessparameter zu schätzen, ist für die praktische Anwendung vorteilhaft: Anstatt eines teuren Messgeräts könnte eine normale Kamera genutzt werden. In den nächsten Abschnitten werden nur die Ergebnisse für das Training des CNNs auf RGB-Bildern diskutiert.

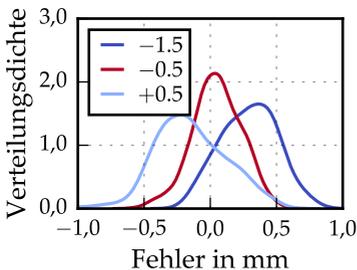
6.3.1.2 Detaillierte Evaluierung der Fehler

Die Verteilung der Fehler auf unbekanntem Daten ist in Abbildung 6.19(a) für die Fokusposition gezeigt. Auf der Abszisse ist die Differenz zwischen Vorhersage und Label aufgetragen. Die zu den Abbildungen gehörenden Mittelwerte und Standardabweichungen sind in Tabelle 6.8 angegeben. Zusätzlich ist vermerkt, wie viele Teile dieser Faktorstufe im Trainingsdatensatz vorlagen. Bei der Fokusposition ist die Verteilung des Fehlers für den LTT-Wert ($-0,5$ mm) symmetrischer um 0 als bei den anderen Fokuspositionen und Standardabweichung sowie MAE sind kleiner. Niedrige Fokuspositionen ($-1,5$ mm) werden tendenziell etwas zu hoch vorhergesagt. Zu hohe Fokuspositionen ($0,5$ mm) werden im Schnitt zu niedrig eingeschätzt. Analog dazu ist die Verteilung des Fehlers für den Vorschub in Abbildung 6.19(b) zu sehen und in Tabelle 6.8 beschrieben. Während der Mittelwert der Fehlerverteilung für 13 und 21 m/min nahe 0 liegt, kommt es bei 29 m/min zu einer deutlichen Abweichung der Vorhersage hin zu niedrigeren Geschwindigkeiten.

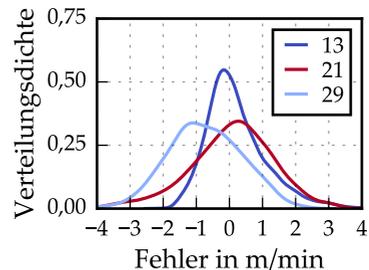
Diskussion Hier macht sich die Unausgeglichenheit des Datensatzes bemerkbar. Aufgrund der Fehlschnitte sind die Faktorstufen im Datensatz ungleich verteilt (siehe Abschnitt 6.4.2), was durch die zufällige Aufteilung in Trainings-, Validierungs- und Testdatensatz noch verstärkt werden

Tabelle 6.7 Mittlerer absoluter Testfehler pro Prozessparameter und Split für drei verschiedene Eingangsdatentypen, ergänzt durch den theoretischen Fehler (der nur von den Labels abhängt).

Daten	Split	ADB in mm	EM in mm	FP in mm	Druck in bar	Vorschub in m/min
RGB	1	0,4	0,3	0,2	1,9	0,9
	2	0,4	0,4	0,2	1,9	1,3
	3	0,3	0,4	0,2	2,0	1,0
	∅	0,4	0,4	0,2	1,9	1,1
Höhe	1	0,4	0,4	0,2	2,3	0,9
	2	0,4	0,5	0,2	2,4	0,8
	3	0,4	0,4	0,2	2,3	1,0
	∅	0,4	0,4	0,2	2,3	0,9
RGBD	1	0,3	0,3	0,2	1,7	1,0
	2	0,4	0,4	0,2	2,1	0,8
	3	0,3	0,4	0,2	1,8	0,8
	∅	0,3	0,4	0,2	1,9	0,9
Theor.	1	0,7	0,9	0,9	3,7	5,1
	2	0,7	0,8	0,8	3,7	4,8
	3	0,7	0,9	0,9	3,6	4,3
	∅	0,7	0,9	0,9	3,7	4,7



(a) Verteilung der Fehler für drei Faktorstufen der Fokusposition.



(b) Verteilung der Fehler für drei Faktorstufen des Vorschubs.

Abbildung 6.19 Verteilung der Fehler (Vorhersage – Label) für verschiedene Labels auf dem Testset 1.

kann. Bei der Fokusposition gilt, dass das Fehlschnittrisiko mit zunehmender Entfernung vom LTT-Wert steigt. Für diesen Wert liegen besonders viele Trainingsbeispiele vor, die Zusammenhänge konnten besser trainiert werden. Der Fehler ist am kleinsten. Es liegt eine kleine Verschiebung hin zu großen Werten vor (Mittelwert 0,1). Dies könnte ebenfalls durch die Datenverteilung im Training verursacht worden sein, da höhere Faktorstufen mit 379 Teilen häufiger vertreten waren als niedrigere Faktorstufen mit 222 Teilen.

Beim Vorschub macht sich die Ungleichverteilung der Trainingsdaten ebenfalls bemerkbar. Allerdings gilt hier, dass das Fehlschnittrisiko für niedrige Geschwindigkeiten besonders klein und für hohe Geschwindigkeiten besonders groß ist. Deshalb sind insbesondere für die Faktorstufe 29 m/min nur wenige Daten vorhanden, was zu einer Verschiebung der Vorhersage hin zu langsamen Werten und zu einem insgesamt größeren Fehler führt. Das Übergewicht der Geschwindigkeiten 13 und 17 m/min führt nicht dazu, dass der LTT-Wert prinzipiell als zu klein eingeschätzt wird.

In Tabelle 6.9 sind die maximalen Fehler des CNNs auf unbekanntem Daten aufgelistet. Beispielsweise beträgt der maximale Fehler für den Vorschub zwischen 5,6 und 7,5 m/min, was 35 bzw. 47 % der Bandbreite und 140 bzw. 188 % der Schrittweite der Parametervariation entspricht. Für die praktische Anwendung müssten die maximalen Fehler deutlich reduziert werden.

6.3.1.3 Vergleich der Trainings-, Validierungs- und Testfehler

In Abbildung 6.20 sind der Trainings- und Validierungsfehler des CNNs für den Split 1 exemplarisch für Vorschub und Gasdruck zu sehen. Das Training wurde nach 64 Epochen beendet, weil sich der Validierungsfehler über 20 Epochen nicht verbessert hatte. Die Modellparameter des CNNs entsprechen denen in Epoche 44. Während der Trainingsfehler in beiden Fällen kontinuierlich abnahm, schwankte der Validierungsfehler von Epoche zu Epoche deutlich. Mit zunehmender Anzahl an Epochen stabilisierte er sich jedoch.

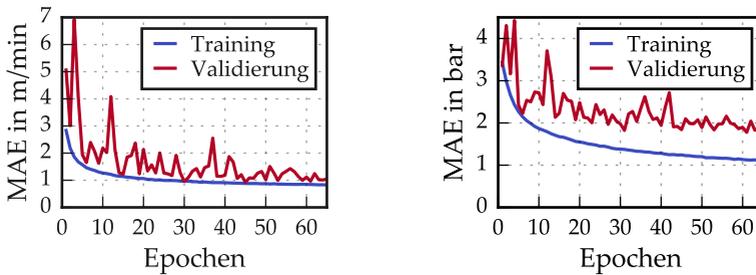
Für den Vorschub (siehe Abbildung 6.20(a)) und die Fokusposition ist die Differenz zwischen Trainings- und Validierungsfehler sehr gering. Zu größeren Unterschieden kommt es jedoch bei Gasdruck (siehe Abbildung

Tabelle 6.8 Anzahl der Teile, Mittelwert und (empirische) Standardabweichung der in 6.19 gezeigten Kurven und MAE für verschiedene Faktorstufen der Fokusposition und des Vorschubs.

Fokusposition	-1,5 mm	-0,5 mm	0,5 mm
Anzahl Teile	52	99	90
Mittelwert	0,3	0,1	-0,1
Standardabweichung	0,3	0,2	0,3
MAE	0,3	0,2	0,3
Vorschub	13 m/min	21 m/min	29 m/min
Anzahl	141	145	63
Mittelwert	0,2	0,1	-0,7
Standardabweichung	0,9	1,3	1,1
MAE	0,7	1,0	1,1

Tabelle 6.9 Vergleich der maximalen Fehler für drei verschiedene Aufteilungen in Trainings-, Validierungs- und Testdaten mit den Bandbreiten und Schrittweiten der Parametervariation.

Daten	Split	ADB in mm	EM in mm	FP in mm	Druck in bar	Vorschub in m/min
Bandbreite	-	2,5	3	5,5	12	16
Schritt	-	0,5	0,5	0,5	3,0	4,0
RGB	1	2,0	1,9	3,2	10,7	5,9
	2	2,0	1,7	2,0	10,2	7,5
	3	1,6	1,7	1,9	11,2	5,6
	∅	1,9	1,8	2,4	10,7	6,3



(a) Mittlerer absoluter Fehler für den Vorschub.

(b) Mittlerer absoluter Fehler für den Gasdruck.

Abbildung 6.20 MAE des Vorschubs und des Gasdrucks über der Anzahl an Epochen für Training und Validierung mit Split 1.

6.20(b)), ADB und EM. In allen drei Fällen sind die Validierungs- und Testfehler um 50 % und mehr größer als die Fehler während des Trainings.

Diskussion Bei ADB, EM und Gasdruck machen sich bereits Überanpassungs-Effekte bemerkbar. Das CNN beginnt, die Parameter zu den Bildern im Trainingsdatensatz „auswendig“ zu lernen, um die Verlustfunktion (den MSE) zu reduzieren. Dieser Effekt ist bei den genannten drei Parametern vermutlich stärker, weil die Zusammenhänge zwischen ihnen und dem Bild der Schnittkante schwieriger zu erkennen sind. Der Trainingsfehler sinkt immer weiter ab, während der Validierungsfehler stagniert. Signifikante Verbesserungen könnten wahrscheinlich durch die Verwendung von mehr Daten erzielt werden.

6.4 Erweiterung des Prozessverständnisses

Wie bereits erläutert wurde (siehe z. B. Abschnitt 1.1.3), verlangen die veränderten Anforderungen an die fertige Industrie eine vollständige Beherrschung des gesamten Produktionsprozesses. Im vorangegangenen Abschnitt wurde bereits ein datenbasiertes Modell, mit dem der Zustand der Laserschneidmaschine überprüft werden kann, vorgestellt. Dies könnte auch dazu beitragen, den Gesamtprozess beherrschbarer zu machen. Im Folgenden wird noch auf zwei weitere Aspekte eingegangen. Zunächst werden die von CNNs erlernten Zusammenhänge zwischen dem Foto der

Schnittkante und den zugrunde liegenden Prozessparametern visualisiert. Dazu wird die Layer-wise Relevance Propagation Methode eingesetzt, die für die Entscheidung des CNNs besonders relevante Pixel im Eingangsbild kennzeichnet. Anschließend wird der Zusammenhang zwischen den Prozessparametern und dem Fehlschnittrisiko, der Rautiefe und der Grathöhe analysiert und modelliert. Für die Modellierung wurden in diesem Fall einfache Regressionsansätze gewählt.

6.4.1 Zusammenhänge zwischen Schnittkante und Prozessparametern

Für die Visualisierung der CNNs wurden zusätzlich zur LRP auch andere Visualisierungsmethoden erprobt, z. B. die Visualisierung einzelner Filter, Regression Activation Maps [Zho+16] und Class Activation Maximization [SVZ14]. Erschwert wird die Visualisierung jedoch durch die große Anzahl und Tiefe der Faltungskerne und Feature Maps, durch die starke Reduktion der räumlichen Auflösung in tiefen Schichten und dadurch, dass Bilder von Schnittkanten für Menschen schlecht zugänglich sind. Außer der LRP lieferte keine der Methoden verständliche Ergebnisse.

6.4.1.1 Klassifikation nach Prozessparametern

Für die Visualisierung mit der LRP wurde für jeden Prozessparameter separat ein CNN trainiert (siehe Abschnitt 5.1.2), mit dem ein RGB-Bild der Schnittkante einer Faktorstufe zugeordnet werden kann. Für den Vorschub unterscheidet das Modell beispielsweise zwischen fünf Klassen: 13, 17, 21, 25 und 29 m/min. In Tabelle 6.10 ist die Anzahl der Klassen jedes Prozessparameters aufgelistet. Außerdem wird dort die gewichtete Korrekturklassifikationsrate der CNN-Klassifikatoren mit der theoretischen Korrekturklassifikationsrate verglichen. Diese würde auftreten, wenn die CNNs im Training lediglich die am häufigsten auftretende Klasse lernen und anschließend beim Testen vorhersagen würden. Wie man in der Tabelle 6.10 sieht, ist die Trefferquote aller Modelle größer als die theoretische. Während die Faktorstufen des Vorschubs und der Fokusposition mit durchschnittlich 95 % bei fünf Klassen und 74 % bei zwölf Klassen sehr gut zugeordnet werden können, sind die Modelle für ADB, EM und Gasdruck deutlich schlechter.

Tabelle 6.10 Anzahl Faktorstufen, gewichtete Korrekturklassifikationsrate (BCA) der fünf CNNs in % verglichen mit der theoretischen Trefferquote.

Daten	Split	ADB	EM	FP	Druck	Vorschub
Klassenanzahl	–	6	7	12	5	5
BCA CNNs in %	1	37	36	76	34	96
	2	37	37	74	46	93
	3	34	42	71	42	96
	∅	36	38	74	41	95
Theor. BCA in %	–	17	14	14	20	20

Diskussion Der Grund dafür ist wahrscheinlich, dass ADB und Gasdruck die Gasströmung verändern und dass ADB und EM die Fokusposition beeinflussen. Auch wenn die Fokusposition richtig erkannt wird, gibt es mehrere Kombinationsmöglichkeiten von ADB und EM. Die Einflüsse der Gasdynamik scheinen nicht genügend zusätzliche Information für die Bestimmung des ADBs bereitzustellen.

Die Trefferquote der Experten für eine ähnliche Aufgabe ist in Abschnitt 6.2.1.2 zu finden. Beim Vergleichen muss man jedoch beachten, dass die befragten Personen lediglich drei Kategorien (größer/kleiner/gleich LTT) unterscheiden mussten.

6.4.1.2 Visualisierung der Zusammenhänge

Nachfolgend werden nur die CNN-Klassifikatoren für Vorschub, Fokusposition und Gasdruck betrachtet. Für jedes gezeigte Beispiel und jeden Klassifikator wurde die LRP durchgeführt und so eine Heatmap berechnet, die die für die Klassenzuordnung wichtigsten Pixel aufzeigt. Anhand von zwei Beispielen kann das Potential der Visualisierung für die Erweiterung des Prozessverständnisses aufgezeigt werden. Das Zeigen aller Aus- und Wechselwirkungen der Prozessparameter in verschiedenen Bereichen des Parameterraums würde den Rahmen dieser Arbeit sprengen.

In Abbildung 6.21 sind die LRP-Heatmaps für zwei verschiedene Schnittkanten gezeigt. Für ein besseres Verständnis der Darstellung soll hier

nochmal kurz darauf eingegangen werden, was die Farben in den Heatmaps bedeuten. Rot steht für einen positiven Beitrag eines Pixels zur Klassifikationsentscheidung des CNNs. Wird ein Bild beispielsweise der Vorschub-Klasse 13 m/min zugeordnet, sind Merkmale der Schnittkante, die für diese Klasse sprechen, rot eingefärbt. Blaue Pixel zeigen dagegen einen negativen Beitrag an. Diese Pixel senken die Wahrscheinlichkeit der Zugehörigkeit zur ausgewählten Klasse (indem sie die Wahrscheinlichkeit der Zugehörigkeit zu einer anderen Klasse erhöhen). Die Ausgänge der CNNs haben einen Wertebereich von $(0, 1]$, die Summe aller Ausgangswerte ergibt 1. Die RGB-Bilder, deren Heatmaps im Folgenden abgebildet sind, werden alle mit einer hohen Wahrscheinlichkeit der korrekten Klasse zugeordnet. Für die negativen Beiträge (Relevanzwerte) verbleiben dadurch nur relativ kleine Werte. Wenn die Klasse 13 m/min beispielsweise mit einer Wahrscheinlichkeit von 0,96 erkannt wird, verbleiben für die anderen vier Faktorstufen nur 0,04. Wichtig zu verstehen ist, dass durch die Skalierung (5.16) nicht darauf geschlossen werden kann, dass ein maximal blau pigmentiertes Pixel denselben Einfluss (mit entgegengesetztem Vorzeichen) wie ein maximal rot pigmentiertes Pixel hatte. Zur Vereinfachung kann man sich auch auf die Betrachtung der roten Pixel beschränken.

Die in Abbildung 6.21 gezeigten Kanten, 0975-1 und 0606-1, wurden mit demselben Vorschub von 13 m/min geschnitten. Da die anderen Prozessparameter verschieden waren, sehen die Kanten unterschiedlich aus. Beim Schneiden von Kante 0975-1 befand sich der Strahlfokus 1,0 mm oberhalb des Blechs, der Gasdruck betrug 9 bar. Bei Kante 0606-1 war der Strahldurchmesser an der Unterkante des 3 mm dicken Blechs am geringsten. Der Gasdruck lag bei 12 bar. Die Bilder beider Kanten werden von drei Klassifikatoren mit einer hohen Wahrscheinlichkeit von jeweils über 92 % korrekt zugeordnet.

Obwohl der Vorschub bei beiden Kanten gleich war, sehen sie unterschiedlich aus. Es gibt nur eine Gemeinsamkeit. Beide Kanten weisen einen eher hohen Grat auf. Kante 0975-1 hat eine Grathöhe von 146 μm , bei Kante 0606-1 ist der Grat sogar 429 μm hoch. Aufgrund dieser Gemeinsamkeit würde man vermuten, dass die Faktorstufe 13 m/min anhand des Grats identifiziert wird. Die LRP-Heatmap 6.21(f) zeigt jedoch, dass diese Eigenschaft für die Klassifikation des Vorschubs irrelevant ist. Hier zeigt sich wieder die Komplexität des Schneidprozesses. Ein langsamer Vorschub

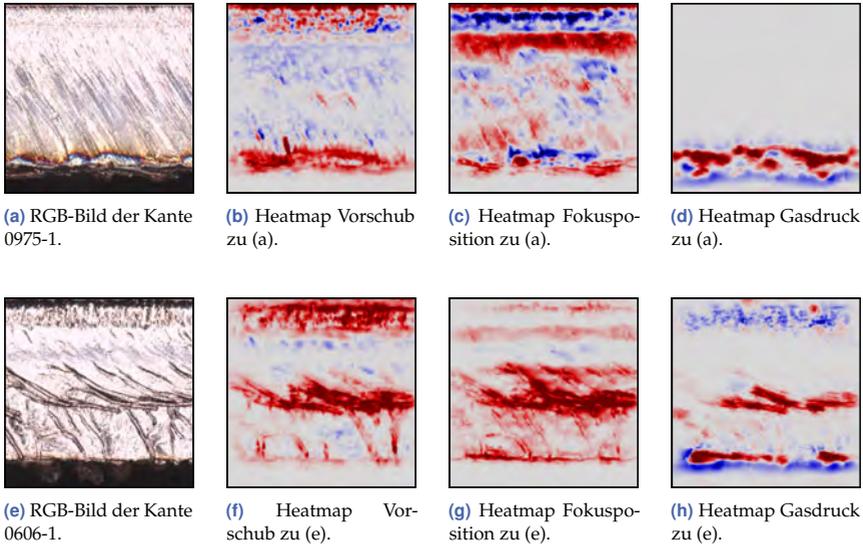


Abbildung 6.21 RGB-Bilder und LRP-Heatmaps der Kanten 0975-1 und 0606-1. Das Teil 0975 wurde mit einem Vorschub von 13 m/min, einer Fokusposition von 1,0 mm und einem Gasdruck von 9 bar geschnitten. Bei Teil 0606 war der Vorschub gleich, die Fokusposition betrug $-3,0$ mm und der Gasdruck 12 bar. In einer Heatmap blau eingefärbte Pixel sind negativ relevant für die Klassifikations-Entscheidung des jeweiligen CNNs, rot eingefärbte Pixel sind positiv relevant (siehe Abschnitt 5.2.2.1).

führt in vielen Fällen zu einem hohen Grat. Die mittlere Grathöhe aller mit 13 m/min geschnittenen Kanten beträgt $207 \mu\text{m}$, im gesamten Datensatz ist sie dagegen nur $104 \mu\text{m}$. Es gibt jedoch auch mit 13 m/min geschnittene Kanten, die nahezu keinen Grat aufweisen. Die Kanten 0338-1, 0637-1 und 0904-1 haben beispielsweise eine Grathöhe von 26, 37 bzw. $27 \mu\text{m}$. Als Mensch würde man bei diesen Kanten wohl falsch liegen. Die Heatmaps könnten dabei helfen, solche Fehlannahmen aufzudecken.

Für den Vorschub-Klassifikator sind die wichtigsten Bereiche bei Kante 0975-1 die Oberkante sowie der verfärbte untere Bereich. Bei Kante 0606-1 sind die Riefenstruktur oben und die Spritzer in der Mitte die wichtigsten Merkmale. Wenn man viele Bilder eines Klassifikators und einer Faktorstufe vergleicht, ist es vielleicht möglich, charakteristische Merkmale abzuleiten.

Neben dem Vergleich der Heatmaps desselben CNNs für verschiedene Kanten ist es auch interessant, die unterschiedlichen Heatmaps einer Kante zu betrachten. Die drei Prozessparameter werden jeweils an anderen Eigenschaften der Schnittkante erkannt. Während der obere Bereich der Schnittfläche von Kante 0975-1 Informationen über Vorschub und Fokusposition enthält, ist er für die Schätzung des Gasdrucks unwichtig. Auch der Riefenverlauf im mittleren Bereich, der für die Fokusposition relevant ist, spielt für den Gasdruck keine Rolle. Nur die verfärbten Pixel im unteren Bereich und die Gratperlen tragen zum Ausgangswert des CNNs bei. Die Tatsache, dass der verfärbte Bereich der Schnittkante den größten positiven Beitrag zur Klassifikation des Gasdrucks leistet, bestätigt, dass es einen starken Zusammenhang mit der Farbe der Schnittkante gibt. Das passt zu dem Ergebnis in Abschnitt 6.3.1.1, dass die Regression des Gasdrucks ohne die Farbinformation schlechter gelingt. Umgekehrt könnte man folgern, dass der Gasdruck hier ausschließlich die Merkmale der Unterkante beeinflusst und dass es deshalb beispielsweise unsinnig wäre diesen Parameter anzupassen, um den Riefenverlauf der Kante zu verändern.

6.4.1.3 Unterstützung bei Schneidproblemen

Indem sie Hinweise liefern, welche Parameteranpassung zielführend sein könnte, können die LRP-Heatmaps also für die Optimierung der Prozessparameter genutzt und als Unterstützung bei Schneidproblemen eingesetzt werden. Dies wird anhand von drei Beispielen gezeigt.

Die Kante des Teils 0987 in Abbildung 6.22(a) weist Verfärbungen und dicke Grattropfen an der Unterkante auf. Wie in der Heatmap in Abbildung 6.22(b) zu sehen ist, besteht ein deutlicher Zusammenhang zwischen der Braunfärbung und dem Vorschub. Die Heatmaps für die Fokusposition in Abbildung 6.22(e) und den Gasdruck in Abbildung 6.22(h) zeigen dagegen keinen klaren positiven Zusammenhang mit den verfärbten Pixeln. Eine Veränderung des Vorschubs von 13 auf 17 bzw. 21 m/min lässt die Verfärbung und die Gratperlen verschwinden, wie in den Abbildungen 6.22(c) bzw. 6.22(d) zu sehen ist. Eine Veränderung der Fokusposition oder des Gasdrucks – egal in welche Richtung – führt dagegen nicht zu einer Verbesserung.

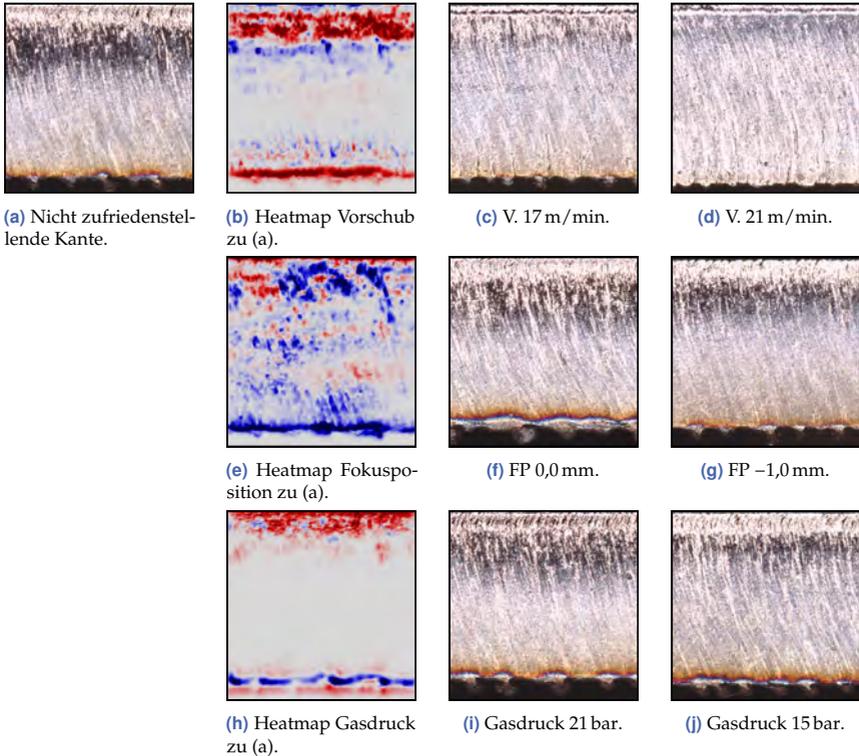


Abbildung 6.22 (a): RGB-Bild von Kante 0987-1 mit ADB 0,5 mm, EM -1,0 mm, FP -0,5 mm, Gasdruck 18 bar, Vorschub 13 m/min. (b): Heatmap zu (a) für den Vorschub. (c) bzw. (d): Verbesserung durch Erhöhung des Vorschubs auf 17 bzw. 21 m/min. (e): Heatmap zu (a) für die Fokusposition. (f) bzw. (g): Veränderung durch höhere bzw. tiefere Fokusposition von 0,0 bzw. -1,0 mm. (h): Heatmap zu (a) für den Vorschub. (i) bzw. (j): Veränderung durch höheren bzw. niedrigeren Gasdruck von 21 bzw. 15 bar.

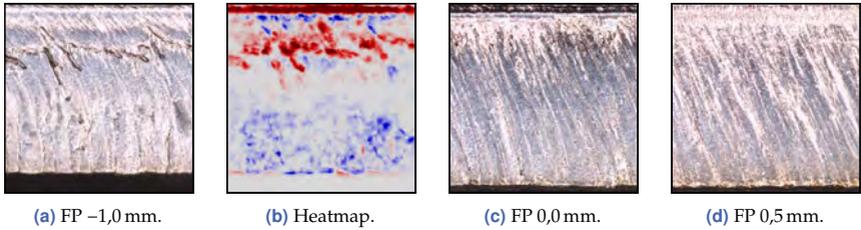


Abbildung 6.23 (a): RGB-Bild von Kante 0882-4 mit ADB 1,5 mm, EM -2,5 mm, FP -1,0 mm, Gasdruck 12 bar. (b): Heatmap zu (a) für die Fokusposition. (c) bzw. (d): Verbesserung durch Erhöhung der Fokusposition auf 0,0 bzw. 0,5 mm.

Dieses Beispiel veranschaulicht, wie die Unterstützung durch die LRP-Heatmaps aussehen könnte. Zunächst muss das unerwünschte Merkmal identifiziert werden. Dann wählt man die Heatmap aus, die den klarsten (positiven!) Zusammenhang damit aufweist. Die Veränderung des zugehörigen Prozessparameters wirkt sich sehr wahrscheinlich auf das störende Merkmal aus. Das verringert die Komplexität der Optimierungsaufgabe erheblich, auch wenn die Heatmaps die Richtung und die erforderliche Schrittweite der Parameteranpassung nicht vorgeben können.

In den Abbildungen 6.23 und 6.24 sind zwei weitere Beispiele in Kurzform gezeigt. Neben dem RGB-Bild ist jeweils lediglich die Heatmap abgebildet, die den eindeutigsten Zusammenhang mit dem störenden Merkmal der Kante aufweist. Rechts daneben sind zwei Kanten gezeigt, die sich bei der Veränderung des zur Heatmap gehörenden Prozessparameters ergeben.

Die Kante des Teils 0882 in Abbildung 6.23(a) könnte verbessert werden, indem die Spritzer in der oberen Hälfte eliminiert werden. Für die Klassifikation der Fokusposition -1,0 mm waren die Spritzer relevant, wie die Heatmap in Abbildung 6.23(b) verdeutlicht. Die Anhebung des Strahlfokus von -1,0 auf 0,0 bzw. 0,5 mm verändert die gesamte Schnittfläche positiv, wie den nebenstehenden Abbildungen 6.23(c) und 6.23(d) zu entnehmen ist. Die Fokusposition wurde hier nur über eine Anpassung des Einstellmaßes verändert. Der ADB blieb unverändert, damit kein Quereinfluss der Gasdynamik auftreten konnte.

Die in der Abbildung 6.24(a) gezeigte Kante des Teils 0663 könnte durch die Entfernung der Spritzer im mittleren Bereich verbessert werden. Da in

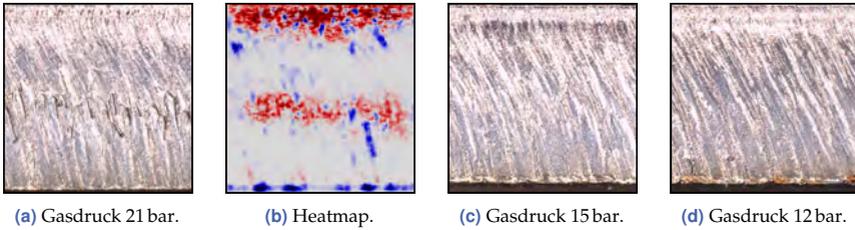


Abbildung 6.24 (a): RGB-Bild von Kante 0663-1 mit ADB 2,0 mm, EM $-1,0$ mm, FP 1,0 mm, Gasdruck 21 bar, Vorschub 21 m/min. (b): Heatmap zu (a) für den Gasdruck. (c) bzw. (d): Verbesserung durch Verringerung des Gasdrucks auf 15 bzw. 12 bar.

der Heatmap des Gasdruck-Klassifikators ein klarer Zusammenhang mit diesem Merkmal besteht (siehe Abbildung 6.24(b)), wurde er von 21 auf 15 bzw. 12 bar verändert. Dadurch verschwinden die Spritzer. Das Resultat ist in den Abbildungen 6.24(c) bzw. 6.24(d) gezeigt.

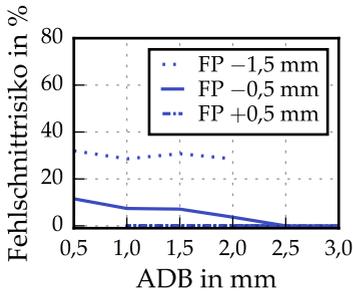
Auch wenn die Heatmaps für eine Verbesserung der Schnittkantenqualität eingesetzt werden können, ist nicht garantiert, dass bei diesem Vorgehen ein Optimum aus Qualität, Stabilität und Produktivität erreicht wird. Nichtsdestotrotz könnten sie als Notfallmaßnahme sehr nützlich sein, wenn es darum geht, überhaupt weiterarbeiten zu können, wenn z. B. während einer Nachtschicht kein Fachpersonal zu erreichen ist.

In den nächsten drei Abschnitten werden die Auswirkungen der Prozessparameter auf die Stabilität des Prozesses sowie auf die Qualität des Schneidergebnisses untersucht und modelliert.

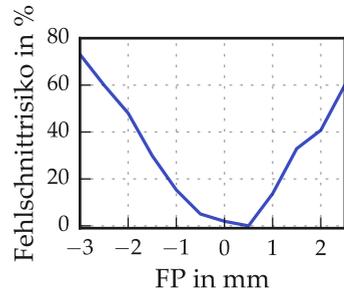
6.4.2 Analyse und Modellierung des Fehlschnittrisikos

Eine verkürzte Fassung der dargestellten Ergebnisse wurde auf der Lasers in Manufacturing Konferenz 2019 vorgestellt [TP19].

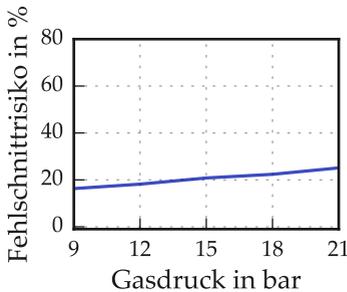
Wie in Abbildung 6.25 für den ADB, die Fokusposition, den Gasdruck und den Vorschub gezeigt ist, wird das Fehlschnittrisiko (siehe Abschnitt 5.3) von den Prozessparametern beeinflusst. Das Einstellmaß fehlt, da es bei der Darstellung von ADB und Fokusposition redundant ist. Um beim ADB den Einfluss auf die Fokusposition vom Einfluss auf die Gasdynamik zu trennen, ist das Fehlschnittrisiko über dem ADB für verschiedene Fokuspositionen aufgetragen. Bei der Interpretation der Zusammenhänge



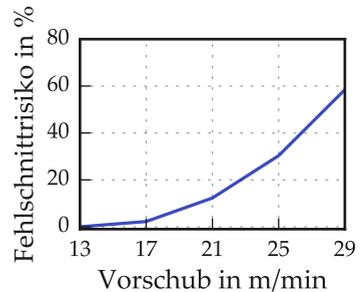
(a) Fehlschnittrisiko in Abhängigkeit des ADBs für konstante Fokuspositionen.



(b) Fehlschnittrisiko in Abhängigkeit der Fokusposition.



(c) Fehlschnittrisiko in Abhängigkeit des Gasdrucks.



(d) Fehlschnittrisiko in Abhängigkeit des Vorschubs.

Abbildung 6.25 Zusammenhänge zwischen dem Fehlschnittrisiko und den Prozessparametern.

sollte man beachten, dass zu jeder auf der Abszisse aufgetragenen Faktorstufe eines Prozessparameters alle Kombinationen der anderen Parameter vorliegen und dass darüber gemittelt wurde. Aufgrund des vollfaktoriellen Versuchsplans liegen zu jeder Faktorstufe eines Prozessparameters dieselben Kombinationen der anderen Parameter vor.

Wie Abbildung 6.25(a) zu entnehmen ist, hängt das Fehlschnittrisiko nur wenig vom Einfluss des ADBs auf die Gasdynamik ab. Es beträgt für tiefe Fokuspositionen über alle ADBs ca. 30%. Bei der Fokusposition 0,5 mm treten keine Fehlschnitte auf. Nur im mittleren Bereich bei einer Fokusposition von -0,5 mm besteht eine geringe Abhängigkeit vom ADB. Je näher sich die Düse am Blech befindet, desto instabiler wird der Prozess.

Die Fokusposition hat einen großen Einfluss auf die Stabilität des Schneidprozesses (siehe Abbildung 6.25(b)). Während bei einer Fokusposition von 0,5 mm, unabhängig von der Wahl der anderen Parameter, kein Fehlschnitt auftritt, beträgt der Anteil der Fehlschnitte bei einer Fokusposition von -3,0 mm 73 % und bei 2,5 mm 60 %. Die Position 0,5 mm über dem Blech scheint hinsichtlich der Prozessstabilität noch besser zu sein als der LTT-Wert (-0,5 mm). Die Empfehlung einer Fokusposition von -0,5 mm lässt sich dadurch erklären, dass sich der Strahlfokus mit zunehmender Erwärmung der Optiken in Richtung der Düse verschiebt (siehe Abschnitt 4.1.3). Da diese Erwärmung üblicherweise nicht kompensiert wird, ist es wichtig, dass dennoch ein stabiler Schneidprozess gewährleistet ist.

Abbildung 6.25(c) zeigt den Einfluss des Gasdrucks auf das Fehlschnittisiko, welches mit steigendem Druck geringfügig ansteigt. Über der gesamten Bandbreite beträgt der Unterschied ca. 9 %. Dieses Ergebnis deckt sich mit dem Einfluss des ADBs, dessen Verringerung vermutlich mit einer Erhöhung des Gasdrucks einhergeht.

Neben der Fokusposition hat der Vorschub den größten Einfluss. Bei einem langsamen Vorschub von 13 m/min wird, unabhängig von der Wahl der anderen Parameter, stets so viel Energie absorbiert, dass das Blech getrennt werden kann. Je höher der Vorschub ist, desto wichtiger wird der Einfluss anderer Parameter und desto wahrscheinlicher wird ein Fehlschnitt. Der Vorschub verändert das pro Zeit zu schmelzende Volumen und wird durch die zur Verfügung stehende Prozessleistung limitiert (siehe Energiebilanz (2.1)). Der LTT-Wert (21 m/min) wurde nicht alleine aufgrund der Stabilität gewählt. Qualität und Produktivität waren auch wichtige Kriterien. Zur Minimierung des Fehlschnitttrisikos sollten eine Fokusposition nahe der Blechoberkante und ein geringer Vorschub verwendet werden. Auch wenn die grundlegenden Auswirkungen eines Prozessparameters so veranschaulicht werden können, sollte man für die Interpretation der Aus- und Wechselwirkungen der Parameter auf ein Modell zurückgreifen.

6.4.2.1 Prädiktion des Fehlschnitttrisikos

Für die Modellierung des Fehlschnitttrisikos wird ein logistischer Regressionsansatz (3.4) verwendet. Das Ziel ist es, die Zugehörigkeit einer Pa-

Tabelle 6.11 Die Wahrheitsmatrix des Modells 4. Grades zur Fehlschnittprädiktion.

		Wahrheit	
		FS	Kein FS
Vorhersage	FS	97 %	5 %
	Kein FS	3 %	95 %

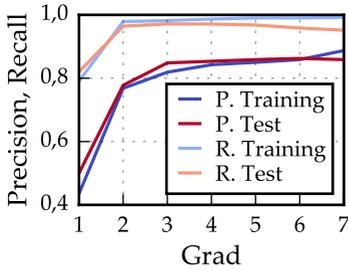
parameterkombination zu einer der beiden Klassen Fehlschnitt (1) oder erfolgreicher Schnitt (0) vorherzusagen. Als Eingangsdaten dienen die Prozessparameter sowie deren polynomiale Erweiterungen (siehe Abschnitt 5.3).

In Abbildung 6.26(a) ist die Modellgüte, beschrieben durch Precision und Recall (siehe Abschnitt 3.1.2), für verschiedene polynomiale Erweiterungen der Eingangsgrößen zu sehen. Die Modellgüte steigt zunächst von Grad 1 zu 2 steil an und verbessert sich dann nur noch langsam. Im Folgenden wird die polynomiale Erweiterung 4. Grades verwendet, da für höhere Grade die Überanpassung zunimmt.

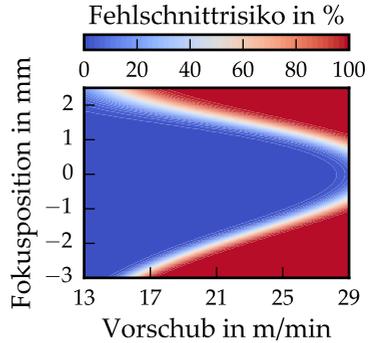
Die Wahrheitsmatrix dieses Modells ist in Tabelle 6.11 zu sehen. 95 % der erfolgreichen Schnitte werden korrekt klassifiziert und 5 % werden als Fehlschnitt eingeordnet. Von den Fehlschnitten werden 97 % korrekt vorhergesagt, 3 % werden fälschlicherweise als erfolgreicher Schnitt eingeschätzt. Der Recall des Modells liegt in allen Fällen über der Precision. Dies ist sinnvoll, da ein guter Schnitt, der fälschlicherweise als Fehlschnitt eingeschätzt wird, deutlich unproblematischer ist als umgekehrt.

Insgesamt hat das Modell 4. Grades 69 erlernte Koeffizienten. Die größten 13 beinhalten jedoch gut die Hälfte der Gesamtrelevanz. Die Geschwindigkeit sowie die Fokuslage sind jeweils in zehn der 13 Eingangsmerkmale mit den größten Koeffizienten enthalten, der Gasdruck und der ABD in fünf bzw. zwei. Die Dominanz des Vorschubs und der Fokuslage deckt sich mit den oben beschriebenen Zusammenhängen.

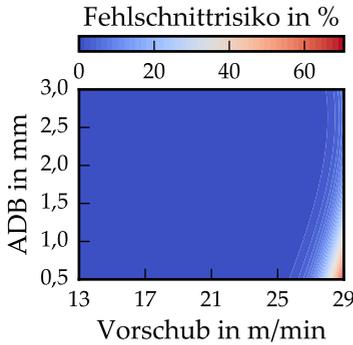
Interpretation Ein solches Modell zur Prädiktion von Fehlschnitten bzw. der Quantifizierung der Prozessstabilität könnte als Grundlage für die Optimierung der Prozessparameter dienen, da es die Grenzen des nutzbaren Parameterraums aufzeigt. In Abbildung 6.26(b) ist das mithilfe



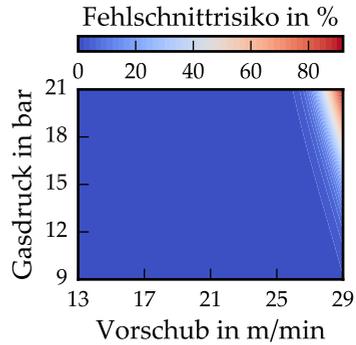
(a) Precision (P.) und Recall (R.) für verschiedene polynomiale Erweiterungen.



(b) ADB 1,5 mm, Gasdruck 15 bar.



(c) ADB 1,5 mm, Fokusposition -0,5 mm.



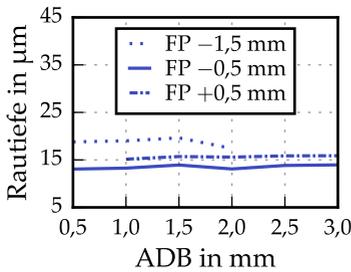
(d) Fokusposition -0,5 mm, Gasdruck 15 bar.

Abbildung 6.26 Güte verschiedener Modelle für die Prädiktion des Fehlschnitttrisikos: (a). Mit dem Modell 4. Grades berechnetes Fehlschnitttrisiko: (b), (c), (d). Es ist zu beachten, dass in den verschiedenen Grafiken unterschiedliche Farbskalen verwendet werden.

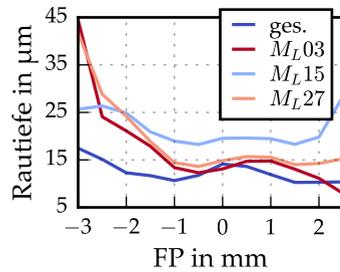
des Modells berechnete Fehlschnitttrisiko über Vorschub und Fokusposition für den LTT-Gasdruck von 15 bar und den LTT-ADB von 1,5 mm zu sehen. Je höher der Vorschub ist, desto kleiner ist die Bandbreite der Fokusposition, innerhalb der das Material getrennt werden kann. Bei der Positionierung des Strahlfokus direkt an der Oberkante kann das Blech mit dem höchstmöglichen Vorschub getrennt werden, weil die einfallende Leistungsdichte maximal ist [WK11]. Ebenfalls gezeigt sind die Abhängigkeiten des Fehlschnitttrisikos von Vorschub und Gasdruck bzw. Vorschub und ADB in den Abbildungen 6.26(c) und 6.26(d). Die anderen Prozessparameter entsprechen dabei den LTT-Werten. Dass der Gasdruck und der Abstand zwischen Düse und Blech die erreichbare Schneidgeschwindigkeit nicht beeinflussen, wurde beispielsweise auch von Borkmann et al. [BMB18] experimentell bestätigt. Man sollte jedoch nicht schlussfolgern, dass die Gasdynamik keinen Einfluss auf die Trennbarkeit des Materials hat. Die Effektivität des Laserschneidprozesses wird maßgeblich durch die Schmelz- und die Materialentfernungsrate beeinflusst [WK10]. Die Fokusposition, welche einen großen Einfluss auf das Fehlschnitttrisiko hat, beeinflusst aber nicht nur die Schmelzrate, sondern sie verändert auch die Breite und Form des Schnittspalts und damit die Gasdynamik und die Materialentfernungsrate [BMB18]. Der Übergangsbereich zwischen einem geringen und einem hohen Fehlschnitttrisiko ist sehr schmal. In diesem Bereich liegen alle Parameterkombinationen aus Tabelle 6.1, mit denen das Blech einmal getrennt werden konnte und einmal nicht.

6.4.3 Analyse und Modellierung der Rautiefe

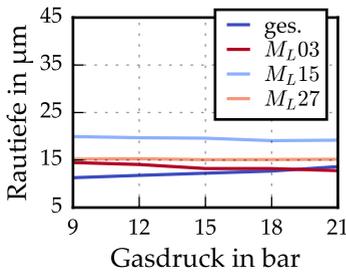
In Abbildung 6.27 ist die Rautiefe für die oberste (03), mittlere (15) und unterste (27) Messlinie sowie gemittelt über alle neun Messlinien (ges.) gezeigt. Über einer Faktorstufe wurde jeweils der Median der Rautiefenwerte aufgetragen. Anders als in Abbildung 6.25 liegen hier – aufgrund der Parameterabhängigkeit der Fehlschnitte – für die verschiedenen Faktorstufen eines Prozessparameters teilweise verschiedene Kombinationen der anderen Parameter vor. Dadurch kann es zu Quereinflüssen kommen. Das ist ein weiterer Grund, warum Modelle bei der Interpretation der Aus- und Wechselwirkungen hilfreich sein können.



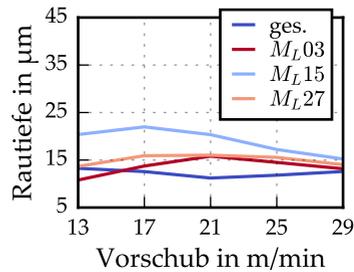
(a) Rautiefe in Abhängigkeit des ADBs für konstante Fokuspositionen.



(b) Rautiefe in Abhängigkeit der Fokusposition.



(c) Rautiefe in Abhängigkeit des Gasdrucks.



(d) Rautiefe in Abhängigkeit des Vorschubs.

Abbildung 6.27 Zusammenhänge zwischen der Rautiefe auf verschiedenen Messlinien und den Prozessparametern.

Wie in den Abbildungen 6.27(a) und 6.27(c) zu sehen ist, beeinflusst eine Veränderung des ADBs und des Gasdrucks die Rautiefe kaum. Es scheint lediglich einen kleinen positiven Einfluss bei der Erhöhung des Gasdrucks zu geben.

Einen großen Einfluss auf die Rautiefe aller Messlinien hat die Fokusposition. Dies ist in Abbildung 6.27(b) gezeigt. Im oberen Bereich der Schnittkante nimmt die Rautiefe mit größer werdender Fokusposition tendenziell ab. Besonders kleine Rz -Werte treten bei $-0,5$ und $2,0$ mm auf. Im mittleren Bereich tritt ein Plateau niedriger Rautiefe für Fokuspositionen zwischen $-1,0$ und $2,0$ mm auf. Außerhalb nimmt die Rautiefe deutlich zu. Im unteren Bereich der Schnittkante wird die Kante bei einer Erhöhung der Fokusposition von $-3,0$ auf $-1,0$ mm glatter. Der Rz -Wert stagniert dann bei ca. $15 \mu\text{m}$. Die durchschnittliche Rautiefe verändert sich weniger

stark als die Werte einzelner Messlinien, deren Verhalten sich im Mittel teilweise kompensiert. Sie ist besonders klein für Fokuspositionen um $-1,0\text{ mm}$ und oberhalb von $1,5\text{ mm}$. Abbildung 6.27(b) zeigt, dass der LTT-Wert von $-0,5\text{ mm}$ ein durchaus sinnvoller Kompromiss für eine mittlere bis niedrige Rautiefe in allen Bereichen ist. Selbst wenn sich der Fokus aufgrund thermischer Effekte von dort aus nach oben bewegt, bleibt die Rautiefe der Schnittkante unverändert.

Der Einfluss des Vorschubs ist in Abbildung 6.27(d) zu sehen. Die Veränderungen der Rautiefe liegen in allen Bereichen bei $5\text{ }\mu\text{m}$ oder weniger über der gesamten Bandbreite.

6.4.3.1 Modellierung der Rautiefe

Der Zusammenhang zwischen den Prozessparametern und der durchschnittlichen Rautiefe wird mittels linearer Regression (3.3) mit polynomialen Kombinationen der Prozessparameter modelliert. Weitere Ergebnisse zur Modellierung der Rautiefe auf den Messlinien 03, 15 bzw. 27 sind im Anhang A.6 zu finden.

In Abbildung 6.28(a) ist die Modellgüte durch den mittleren absoluten Fehler für verschiedene polynomiale Erweiterungen jeweils separat für die Trainings- und Testdaten beschrieben. Die Entscheidung fiel hier, wie bei der Modellierung des Fehlschnittrisikos, auf die polynomiale Erweiterung 4. Grades, da diese einen guten Kompromiss aus höchstmöglicher Güte und kleinstmöglicher Überanpassung darstellt. Der mittlere absolute Fehler beträgt $1,6\text{ }\mu\text{m}$ im Training und $1,7\text{ }\mu\text{m}$ für unbekannte Parameterkombinationen. Der (theoretische) Fehler eines Modells, welches im Test stets den Mittelwert der Rautiefe der Trainingsdaten vorhersagt, wäre $3\text{ }\mu\text{m}$.

Insgesamt hat das Modell 4. Grades 69 erlernte Koeffizienten. Während bei der Prädiktion des Fehlschnittrisikos die 13 größten Koeffizienten gleichviel Einfluss haben wie die 56 kleinsten, sind es hier die 20 größten bzw. 49 kleinsten. Die Fokusposition ist in 16 dieser 20 Eingangsvariablen enthalten, der Vorschub in elf, der Gasdruck in zehn und der ADB in drei. Hier ist die Fokusposition die dominierende Einflussgröße.

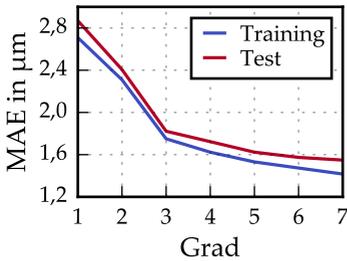
Interpretation In Abbildung 6.28(b) ist die vom Modell geschätzte Rautiefe in Abhängigkeit des Vorschubs und der Fokusposition zu sehen. Die

nicht variierten Prozessparameter entsprechen jeweils dem LTT-Wert. Der Fokusbereich, in dem kleine Rautiefenwerte erzielt werden, wird mit zunehmender Schneidgeschwindigkeit etwas kleiner. Die Zunahme der Rautiefe in Abhängigkeit der Fokusposition bei einer festen Geschwindigkeit verhält sich ungefähr symmetrisch zur Fokusposition 0. Dies entspricht einem Strahlfokus an der Oberkante des Blechs. Besonders glatte Kanten sind bei einem niedrigen Vorschub und einer Fokusposition 1,0 mm unter- und oberhalb der Blechoberfläche möglich. Dass die Rautiefe besonders niedrig ist, wenn sich der Laserstrahl langsam bewegt, könnte daran liegen, dass in diesem Fall ein sehr breiter Schnittpalt entsteht, welcher die Aufrechterhaltung einer laminaren Strömung ermöglicht, was sich glättend auswirkt [WK10]. Die Rautiefe ist groß, wenn der Strahlfokus tief im Blech positioniert ist. Es ist jedoch zu beachten, dass das mögliche Prozessfenster für diese Fokuspositionen mit steigendem Vorschub kleiner wird und Teile mit einer Fokusposition unter $-2,0$ mm nicht schneller geschnitten werden können als mit 21 m/min. Die Abbildung 6.28(b) sollte deshalb stets nur unter Zuhilfenahme von Abbildung 6.26(b) interpretiert werden. Wie in den Abbildungen 6.28(c) und 6.28(d) zu sehen ist, besteht nur ein geringer Zusammenhang zwischen ADB bzw. Gasdruck und Rautiefe. Es bestätigt sich aber der Zusammenhang, dass sich ein größerer Gasdruck oder ein kleinerer ADB positiv auf die Rautiefe auswirken, weil dadurch vermutlich die Aufrechterhaltung der laminaren Strömung im gesamten Schnittpalt begünstigt wird.

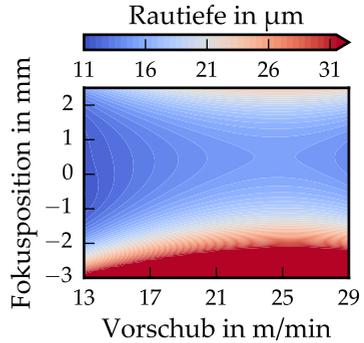
6.4.4 Analyse und Modellierung des Grats

In Abbildung 6.29 sind die Zusammenhänge zwischen der Grathöhe und den Prozessparametern visualisiert. Über einer Faktorstufe wurde jeweils der Median der Grathöhenwerte aufgetragen.

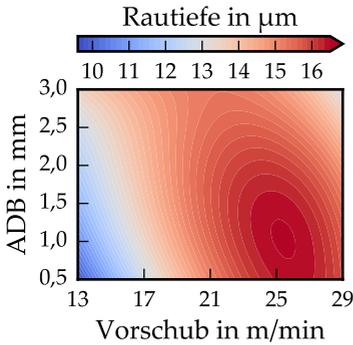
Der ADB hat, abhängig von der Fokusposition, keinen oder einen großen Einfluss auf die Grathöhe. Je tiefer im Blech der Strahlfokus positioniert ist, desto größer ist der Einfluss der Gasdynamik, die durch den ADB verändert wird. Mit kleiner werdendem ADB wächst der Grat. Es ist davon auszugehen, dass sich der Druck im Schnittpalt erhöht, wenn sich die Düse näher an die Blechoberseite befindet. Wie man in Abbildung 6.29(c) sieht, sorgt auch eine Erhöhung des Gasdrucks für eine zunehmende Grathöhe.



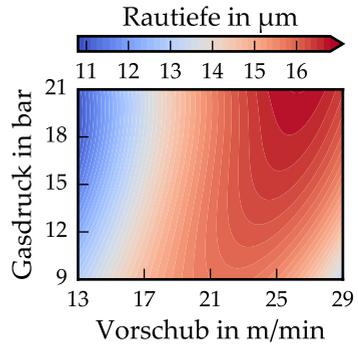
(a) Mittlerer absoluter Fehler für verschiedene polynomiale Erweiterungen.



(b) ADB 1,5 mm, Gasdruck 15 bar.

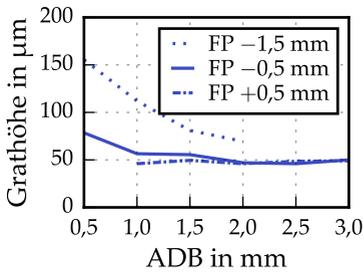


(c) Fokusposition $-0,5$ mm, Gasdruck 15 bar.

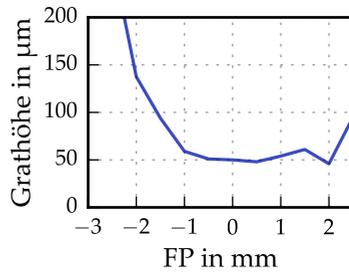


(d) ADB 1,5 mm, Fokusposition $-0,5$ mm.

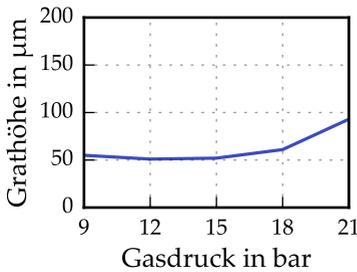
Abbildung 6.28 Güte verschiedener Modelle für die Prädiktion der Rautiefe: (a). Mit dem Modell 4. Grades berechnete Rautiefe: (b), (c), (d). Es ist zu beachten, dass in den verschiedenen Grafiken unterschiedliche Farbskalen verwendet werden.



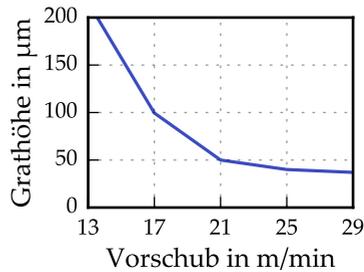
(a) Grathöhe in Abhängigkeit des ADBs für konstante Fokuspositionen.



(b) Grathöhe in Abhängigkeit der Fokusposition.



(c) Grathöhe in Abhängigkeit des Gasdrucks.



(d) Grathöhe in Abhängigkeit des Vorschubs.

Abbildung 6.29 Zusammenhänge zwischen der Grathöhe und den Prozessparametern.

Die Fokusposition ist wie beim Fehlschnittisiko und bei der Rautiefe eine entscheidende Einflussgröße. Insbesondere sehr tiefe Fokuspositionen verursachen sehr hohe Grate. Von $-1,0$ bis $1,0$ mm verändert sich die Grathöhe kaum. Die Ausbildung eines Plateaus in diesem Bereich war auch bei der Rautiefe zu beobachten.

Der Vorschub hat ebenfalls einen großen Einfluss auf die Gratbildung. Die Grathöhe nimmt steil ab bei einer Erhöhung des Vorschubs von 13 auf 21 m/min.

Die Gegenüberstellung der Abbildungen 6.25(d) und 6.29(d) zeigt, wie schwierig die Optimierung der Prozessparameter ist. Für ein geringes Fehlschnittisiko sollte tendenziell ein niedriger Vorschub gewählt werden. Dieser verursacht jedoch eine starke Gratbildung.

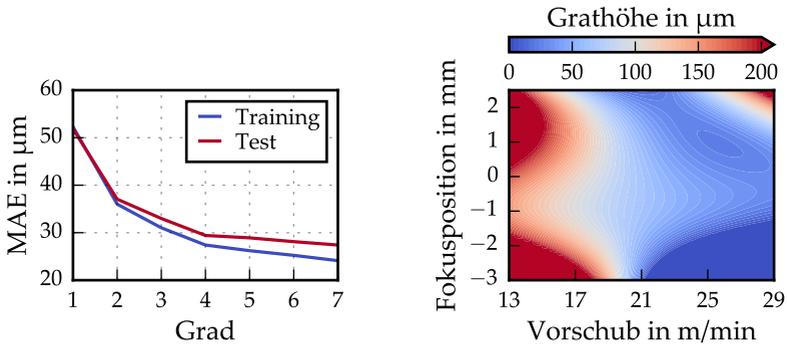
6.4.4.1 Modellierung der Grathöhe

Der Zusammenhang zwischen den Prozessparametern und der Grathöhe wird wieder mittels linearer Regression abgebildet. Polynomiale Kombinationen der Prozessparameter werden als Eingangsvariablen verwendet.

In Abbildung 6.30(a) ist die Modellgüte beschrieben durch den mittleren absoluten Fehler für verschiedene polynomiale Erweiterungen jeweils separat für Trainings- und Testdaten ausgewertet. Auch hier fiel die Entscheidung auf die polynomiale Erweiterung 4. Grades. Der MAE beträgt $27\ \mu\text{m}$ im Training und $30\ \mu\text{m}$ für unbekannte Parameterkombinationen. Der Fehler eines Modells, welches im Test stets den Grathöhenmittelwert der Trainingsdaten vorhersagt, wäre $76\ \mu\text{m}$.

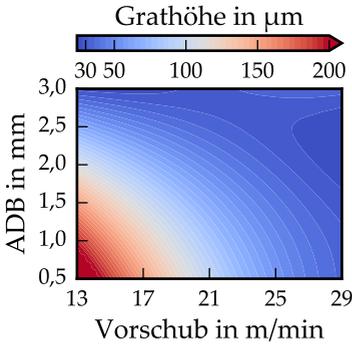
Insgesamt hat das Modell 69 erlernte Koeffizienten. 17 Koeffizienten beinhalten die Hälfte der Gesamtrelevanz. Die Fokuslage ist in zwölf dieser 17 Variablen enthalten, der Vorschub in acht, der Gasdruck in sieben und der ABD in vier. Hier stellt auch die Gasdynamik einen wichtigen Einfluss dar.

Interpretation Die Zusammenhänge zwischen der vom Modell vorhergesagten Grathöhe in Abhängigkeit des Vorschubs und der Fokusposition für einen ADB von $1,5\ \text{mm}$ und einen Gasdruck von $15\ \text{bar}$ sind in Abbildung 6.30(b) zu sehen. Da dieses Modell die Grenzen des Prozessbereichs nicht kennt, sollte diese Abbildung nur gemeinsam mit Abbildung 6.26(b) verwendet werden. Beim Vergleich dieser beiden Abbildungen fällt auf, dass kleine Grathöhen vor allem an der Grenze zum nicht schneidbaren Bereich auftreten. Bei hohen Schneidgeschwindigkeiten ist der Wärmeverlust in das umgebende Material besonders klein. Dies führt zu einer höheren Temperatur und damit zu einer besseren Fließfähigkeit der Schmelze, vereinfacht die Materialentfernung und führt dadurch wahrscheinlich zu einem kleineren Grat [WK10]. Bei einem niedrigen Vorschub ist der Grat dagegen hoch. Besonders große Gratperlen entstehen bei einem niedrigen Vorschub in Kombination mit einer ungünstigen Fokusposition von über $1,0$ oder unter $-2,0\ \text{mm}$. In den Abbildungen 6.30(c) und 6.30(d) zeigt sich ein gegenüber der Rautiefe gegensätzliches Verhalten. Während ein niedriger ADB und ein hoher Gasdruck bei geringen Geschwindigkeiten zu einer verminderten Rautiefe führen (siehe Abbildung 6.28(c) bzw. 6.28(d)), kommt es zu einer Zunahme der Grathöhe. Auch hier muss also

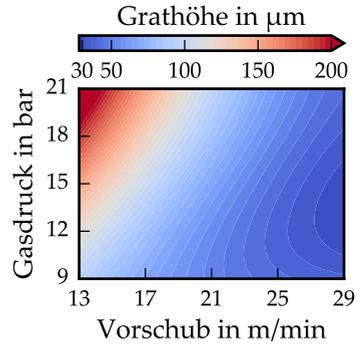


(a) Mittlerer absoluter Fehler für verschiedene polynomiale Erweiterungen.

(b) ADB 1,5 mm, Gasdruck 15 bar.



(c) Fokusposition $-0,5$ mm, Gasdruck 15 bar.



(d) ADB 1,5 mm, Fokusposition $-0,5$ mm.

Abbildung 6.30 Güte verschiedener Modelle für die Prädiktion der Grathöhe: (a). Mit dem Modell 4. Grades berechnete Grathöhe: (b), (c), (d).



(a) Optimierung von Stabilität und Rautiefe. Kante 0546-1.

(b) Minimierung der Grathöhe. Kante 0838-1.

Abbildung 6.31 Exemplarische Kanten für unterschiedliche Präferenzen.

ein Kompromiss zwischen verschiedenen Qualitätsmerkmalen gefunden werden. Allerdings verursachen ADB und Gasdruck beim Grat Änderungen im dreistelligen μm -Bereich, während bei der Rautiefe nur kleine Unterschiede auftreten.

6.4.5 Praktische Anwendung der Modelle

Mit den oben beschriebenen Modellen für die Prozessstabilität, die mittlere Rautiefe und die Grathöhe wäre es möglich, die Prozessparameter an individuelle Kundenwünsche anzupassen. Die Präferenzen hinsichtlich verschiedener Prozessmerkmale könnten genutzt werden, um die Modelle zu gewichten. Basierend auf der gewichteten (und normierten) Summe der Modelle, könnte eine individuell optimale Prozessparameterkombination berechnet werden. Das Modell für die Prädiktion des Fehlschnitttrisikos würde die Randbedingungen für diese Optimierung vorgeben.

Bei gleicher Gewichtung von Stabilität und Rautiefe würde die Kante beispielsweise wie die in Abbildung 6.31(a) gezeigte aussehen. Diese wurde mit einem Vorschub von 13 m/min und einer Fokusposition von $-0,5$ mm geschnitten. ADB und Gasdruck entsprachen den LTT-Werten. Die Kante weist allerdings einen deutlichen Grat auf. Soll nur die Grathöhe optimiert werden, würde sich z. B. die in Abbildung 6.31(b) gezeigte Kante ergeben. Hier betrug der Vorschub 25 m/min, die Fokusposition 1,0 mm. ADB und Gasdruck entsprachen ebenfalls LTT-Werten. In diesem Fall befindet man sich allerdings bereits nahe an der Grenze zum nicht schneidbaren Bereich (siehe Abbildung 6.26(b)). Diverse andere Gewichtungen der Merkmale sind denkbar. Zusätzlich sollten weitere Qualitätskriterien einbezogen werden, z. B. die Homogenität und die Verfärbung (siehe Abschnitt 6.2.1).

Die visualisierten Zusammenhänge könnten außerdem bei der manuellen Ermittlung geeigneter Prozessparameter helfen, indem sie erstens aufzeigen, mit welchem Parameter die größten Veränderungen erzielt werden können und zweitens die Richtung vorgeben, wie welcher Parameter je nach Präferenz eingestellt werden sollte.

7 Schluss

In Kapitel 1 der vorliegenden Arbeit wurden drei Anforderungen an die fertigende Industrie formuliert: Eingehen auf individuelle Kundenwünsche, sparsame Nutzung von Ressourcen und günstige Stückkosten auch für kleine Losgrößen. Obwohl das Laserschneiden aufgrund seiner Flexibilität gut geeignet wäre, diesen Anforderungen zu begegnen, ist man ihnen mit der heutigen Arbeitsweise nicht gewachsen. Alle erarbeiteten Ergebnisse zielen deshalb auf eine Verbesserung eines oder mehrerer der genannten drei Punkte ab. Aufgrund des großen Erfolgs der Methoden des maschinellen Lernens in verschiedensten Bereichen und weil diese Methoden bislang nur wenig Anwendung im Laserschneiden finden, wurden in dieser Arbeit ausschließlich datenbasierte Modellansätze verwendet. Es ging dabei auch darum, das Potential des maschinellen Lernens für Anwendungen im Laserschneiden aufzuzeigen.

Da zu Beginn der Arbeit keine Daten vorhanden waren, musste zunächst eine breite Datenbasis erzeugt werden. Es wurden 1126 Teile mit verschiedenen Prozessparameterkombinationen geschnitten. Sofern bei einem Teil kein Fehlschnitt auftrat und es aus dem Restgitter entnommen werden konnte, wurden von jeder Schnittkante RGB-Bilder und 3D-Punktwolken aufgenommen. Aus den 3D-Punktwolken konnten objektive Qualitätsmerkmale wie Rautiefe und Grathöhe abgeleitet werden. Ergänzt wurden diese mit den subjektiven Qualitätsbewertungen durch Experten. Die Qualität der Datenbasis wurde analysiert und es wurden verschiedene Methoden zur Korrektur von Rautiefenmessfehlern evaluiert. Die größte Verbesserung der Datenqualität konnte durch zweimaliges Vermessen jeder Kante erzielt werden. Fast genauso gute Resultate lieferte die Korrektur der Messdaten eines Durchgangs mit einem Autoencoder.

Die durchgeführte Expertenbefragung zeigte, dass kein allgemein gültiges Qualitätsverständnis angenommen werden kann und man stattdessen auf individuelle Präferenzen eingehen sollte, weil die Bewertung derselben Schnittkante durch verschiedene Personen teils sehr unterschiedlich

ausfiel. Die Rauheit, die Homogenität und die Grathöhe waren für alle bewertenden Personen wichtig. Die Verfärbung der Schnittkante war dagegen für manche Personen relevant und für andere nicht. Interessant war, dass viele als gut bewertete Schnittkanten mit gegenüber dem LTT-Wert erhöhten Vorschüben geschnitten wurden. Bei einer verbesserten Kontrollierbarkeit des Gesamtprozesses und der Maschine könnte man Sicherheitsfaktoren verkleinern, schneller schneiden und so eine bessere Qualität und gleichzeitig eine höhere Produktivität erreichen. Bei der Expertenbefragung zeigte sich außerdem, dass Menschen generell Schwierigkeiten haben, die Qualität der Schnittfläche einer Kante zu beurteilen. Deshalb sollte die subjektive Beurteilung durch objektive Kenngrößen ergänzt werden. Zu diesem Zweck wurde ein faltendes neuronales Netz entwickelt, mit dem die Rautiefe einer Schnittkante basierend auf einem RGB-Bild geschätzt werden kann, damit sie auch ohne ein teures Messgerät zur Verfügung steht. Es zeigte sich, dass es prinzipiell möglich ist, den Zusammenhang zwischen Bild und Rautiefe zu erlernen. Gleichzeitig konnte an diesem Beispiel gezeigt werden, wie wichtig die Qualität der Trainingsdaten für die Modellgüte ist. Das auf den unbereinigten Messdaten trainierte CNN schätzte die Rautiefe systematisch schlechter ein, als das auf den (mit dem Autoencoder) korrigierten Daten trainierte Netz.

Ein CNN eröffnet außerdem eine neue Möglichkeit, den Zustand einer Laserschneidmaschine zu überprüfen. Mithilfe des Netzes können die einer Schnittkante zugrunde liegenden Prozessparameter ausgehend von einem RGB-Bild geschätzt werden. Besteht eine Differenz zwischen diesen Ist-Prozessparametern und den eingestellten Soll-Parametern, liegt ein Problem in der Maschine vor. Die Kenntnis des betroffenen Parameters verkleinert den Suchraum erheblich und verkürzt so die Arbeits- und die Maschinenstillstandszeit. Dadurch können teure Fachkräfte effizienter eingesetzt und die Maschinen produktiver betrieben werden. Da voll automatisierte Laserschneidmaschinen mit dieser Methode dauerhaft überwacht werden können, kann die Maschine bei Problemen angehalten und eine aufwendige Nachbearbeitung vermieden werden. Die Bestimmung der Schneidparameter aus dem RGB-Bild gelang besonders gut für Vorschub und Fokuspotion und etwas weniger gut für den Abstand zwischen Düse und Blech, das Einstellmaß und den Gasdruck. Da Menschen angesichts der großen Komplexität des Prozesses nicht in der Lage sind,

von der Schnittkante auf die Prozessparameter zu schließen, zeichnete sich das große Potential der Methoden des maschinellen Lernens für das Laserschneiden hier besonders deutlich ab.

CNNs, die ein Bild der Schnittkante der Faktorstufe eines Prozessparameters zuordnen, können zusammen mit einer geeigneten Visualisierungsmethode dazu genutzt werden, das Prozessverständnis zu erweitern. Das ist eine wichtige Voraussetzung für die Herstellung kleiner Losgrößen zu günstigen Preisen. Mit der Layer-wise Relevance Propagation können die für die Entscheidung des Netzes wichtigsten Pixel des Eingangsbildes in einer Heatmap deutlich gemacht werden. So wird für den Menschen verständlicher, was innerhalb des neuronalen Netzes passiert. Die Heatmaps können auch als erste Hilfe bei Schneidproblemen genutzt werden. Treten beispielsweise unerwünschte Spritzer auf der Kante auf und zeigt die Visualisierung an, dass dieses Merkmal besonders wichtig für die Erkennung des Gasdrucks war, folgt umgekehrt, dass der Gasdruck dieses Merkmal besonders stark beeinflusst und die Spritzer verschwinden, wenn er geändert wird.

Ein weiterer Beitrag zur Erweiterung des Prozessverständnisses sind die detaillierten Analysen und die Modellierung der Zusammenhänge zwischen den Prozessparametern und den drei Merkmalen Fehlschnittisiko, Rautiefe und Grathöhe. Die Modellierung dieser Zusammenhänge auf dem Fundament der breiten Datenbasis, die alle Wechselwirkungen der variierten Prozessparameter beinhaltet, bietet gleich mehrere Anwendungsmöglichkeiten. Die Modelle zeigen, wie schwierig die Optimierung des Laserschneidprozesses ist, weil stets ein Kompromiss gefunden werden muss zwischen Prozessstabilität, verschiedenen Qualitätsmerkmalen und der Produktivität. Beispielsweise wird die Grathöhe an der Grenze des schneidbaren Parameterraums minimal. Die Modelle können bei der manuellen Suche nach einem geeigneten Kompromiss unterstützen, weil sie alle Aus- und Wechselwirkungen der Parameter transparent machen. Eine weitere Einsatzmöglichkeit ist die individuelle Optimierung des Schneidprozesses. Die Modelle ermöglichen die Berechnung einer personalisierten Prozessparameterkombination basierend auf der individuellen Gewichtung der Prozess- und Qualitätsmerkmale ohne eine aufwendige experimentelle Anpassung. Die Kosten wären dann unabhängig von der Stückzahl.

7.1 Ausblick

Für eine sinnvolle multikriterielle Optimierung des Schneidprozesses sollten die Zielfunktionen Fehlschnitttrisiko, Rautiefe und Grathöhe erweitert werden durch eine Größe, die die vom Menschen wahrgenommene Homogenität quantifiziert sowie durch die Verfärbung und die Rechtwinkligkeitstoleranz.

Um eine Vorstellung der tatsächlichen Bandbreite bei der Qualitätsbewertung von Schnittkanten zu erhalten, sollte die Befragung auf Personen unterschiedlicher Branchen in verschiedenen Märkten ausgeweitet werden. Außerdem wäre es sinnvoll, den vorhandenen Maschinenpark zur Nach- und Weiterbearbeitung einzubeziehen.

Alle in der vorliegenden Arbeit präsentierten Ergebnisse wurden auf derselben Datenbasis entwickelt. Diese besteht nur aus Edelstahl-Schnittkanten der Blechdicke 3 mm, die auf einer Maschine mit verschiedenen Kombinationen von vier Prozessparametern geschnitten wurden. Die wichtigsten nächsten Schritte sind deshalb die Untersuchung der Übertragbarkeit auf andere Blechdicken, andere Materialien und andere Maschinen sowie die Einbeziehung zusätzlicher Prozessparameter. Die Hinzunahme jeder weiteren Einflussgröße erhöht jedoch die Komplexität und vergrößert die erforderliche Datenmenge. Des Weiteren sollte die Datenqualität, insbesondere bei der Erfassung der 3D-Punktwolken, verbessert werden.

Da diese Arbeit nicht eine Nische eines bereits gut erforschten Gebietes behandelt, sondern ein neues Anwendungsgebiet datenbasierter Methoden eröffnet, ergibt sich daraus eine Vielzahl an weiteren Fragestellungen, die weit über die bisherige Datenbasis und die beschriebenen Erweiterungen hinausgehen.

Anhang

A Anhang

A.1 Technische Daten

In diesem Abschnitt sind die technischen Daten der verwendeten Laserschneidmaschine und der Strahlquelle (siehe Tabelle A.1) sowie des Keyence Messgeräts VR-3200 (siehe Tabelle A.2) aufgelistet. Außerdem werden Angaben zur Hardware (siehe Tabelle A.3) und den eingesetzten Python-Bibliotheken gemacht (siehe Tabelle A.4).

Tabelle A.1 Technische Daten der Laserschneidmaschine TruLaser 5030 fiber (L68) und der Strahlquelle TruDisk 12001 von TRUMPF. Quelle: TRUMPF.

Arbeitsbereich x	3000 mm
Arbeitsbereich y	1500 mm
Maximale Geschwindigkeit	280 m/min
Maximale Blechdicke Edelstahl	40 mm
Wellenlänge	1030 nm
Strahlparameterprodukt (vor Lichtleitkabel)	4 mm mrad
Numerische Apertur (nach Lichtleitkabel)	0,1
Kontinuierlich einstellbarer Leistungsbereich	240 W bis 12 kW

Tabelle A.2 Technische Daten des Messgeräts VR-3200 der Keyence Corporation im Modus 40-fache Vergrößerung. Quelle: Keyence Corporation.

Betrachtungsmessbereich	7,6 mm × 5,7 mm
Messbare Höhe	±0,5 mm
Wiederholbarkeit	0,5 µm (Höhe), 0,5 µm (Breite)
Messgenauigkeit	±3 µm (Höhe), ±2 µm (Breite)
XY-Hub Messtisch	184 mm × 88 mm
Z-Hub Messtisch	90 mm
Bildaufnahmelement	4 Mio. Pixel CMOS, monochrom
Projektionsobjektive	Doppelt telezentrisch
Lichtempfangendes Objektiv	Doppelt telezentrisch
Lichtquelle für Betrachtung	Ringleuchte (rot/grün/blau)
Lichtquelle für Messung	Weißer LED

Tabelle A.3 Verwendete Hardware zum Training der ML-Modelle.

Prozessor	Intel® Core™ i9 Ten, i9-9820X
Arbeitsspeicher	64 GB-Corsair VENGEANCE DDR4 3000 MHz
Grafikkarte	GB-NVIDIA GeForce RTX 2080 Ti

Tabelle A.4 Wichtige verwendete Python-Bibliotheken mit Versionsnummer.

Bibliothek	Version
Python	3.7.3
Keras [Cho15]	2.2.4
NumPy [Har+20]	1.16.2
Matplotlib [Hun07]	3.0.3
pandas [McK10]	0.24.2
OpenCV [Bra00]	3.4.1
scikit-learn [Ped+11]	0.20.3
TensorFlow [Aba+16]	1.13.1

A.2 LTT-Teile

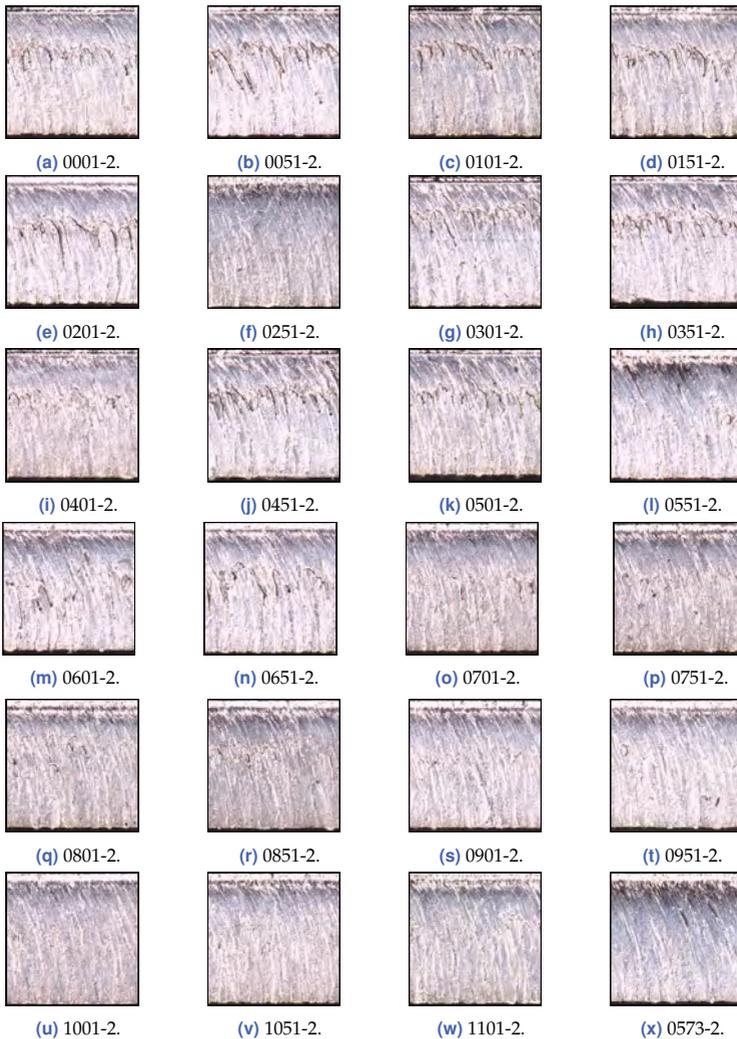


Abbildung A.1 RGB-Bilder der Schnittfläche von Kante 2 aller LTT-Teile. Gezeigt ist jeweils ein ca. 3 mm × 3 mm großer Ausschnitt.

A.3 Schnittkanten der Teile 0949, 0992, 1006, 1056, 1077



Abbildung A.2 Ergänzung zu Tabelle 6.6: RGB-Bildausschnitte der Teile 0949, 0992, 1006, 1056 und 1077 jeweils mit den Kanten -1, -2, -3 und -4. Gezeigt ist jeweils ein ca. 3 mm × 3 mm großer Ausschnitt.

A.4 LRP-Regeln

Dieser Abschnitt schließt an die Herleitung der LRP über die Deep Taylor Decomposition in Abschnitt 5.2.1 an.

Es gibt verschiedene Regeln für die Rückpropagation der Relevanz. Diese haben gegenüber Formel (5.12) beispielsweise vorteilhafte numerische Eigenschaften (ϵ/γ -Regel) oder sie berücksichtigen, dass nicht beliebige reelle Zahlen als Eingangswerte einer Schicht zugelassen sind ($0/z^\beta$ -Regel). Dafür müssen die Forderungen nach Erhaltung (5.1) und Positivität (5.2) teilweise gelockert werden.

LRP-0-Regel: Die LRP-0-Regel ergibt sich, wenn nicht alle reellen Zahlen zugelassen sind und im Intervall $(0, \{a_i\})$ nach der Entwicklungsstelle gesucht wird. Gesucht wird dabei entlang des Vektors selbst (und nicht entlang des Gradienten) in Richtung des Ursprungs. Es gilt [Bac+15]:

$$R_i = \sum_j \frac{z_{ij}}{\sum_i z'_{ij}} \cdot R_j \text{ mit } z_{ij} = a_i w_{ij}. \quad (\text{A.1})$$

Auch wenn diese Regel auf den ersten Blick sinnvoll scheint, hat sich gezeigt, dass ihre Anwendung auf das gesamte neuronale Netz der Multiplikation des Gradienten mit dem Eingang entspricht [SGK17]. Der Gradient tiefer neuronaler Netze ist jedoch üblicherweise verrauscht [Bal+17], was zu schwierig interpretierbaren Heatmaps führt. Die LRP-0-Regel ist außerdem numerisch instabil. Bach et al. [Bac+15] und Montavon et al. [Mon+17; Mon+19] haben deshalb modifizierte Regeln mit besseren numerischen Eigenschaften entwickelt.

LRP- ϵ -Regel: Ergänzt man ϵ im Nenner der LRP-0-Regel, ergibt sich der folgende Zusammenhang [Bac+15]:

$$R_i = \sum_j \frac{z_{ij}}{\epsilon + \sum_i z'_{ij}} \cdot R_j. \quad (\text{A.2})$$

Die Ergänzung von ϵ sorgt für numerische Stabilität. Die resultierenden Heatmaps sind weniger verrauscht, die Relevanz einiger Pixel verschwindet jedoch, weil ϵ kleine Beiträge absorbiert. In dieser Arbeit wurde $\epsilon = 0,25$ verwendet.

LRP- γ -Regel: Diese Regel entspricht der LRP-0-Regel, aber z_{ij} verändert sich zu [Mon+19]:

$$z_{ij} = a_i \cdot (w_{ij} + \gamma \cdot w_{ij}^+). \quad (\text{A.3})$$

Dadurch werden die positiven Beiträge w_{ij}^+ gegenüber den negativen bevorzugt. Diese Bevorzugung fällt umso stärker aus, je größer γ ist. Die Übergewichtung der positiven Beiträge erhöht die Stabilität, kann jedoch dazu führen, dass eher unwichtige Merkmale auch als relevant angezeigt werden. In dieser Arbeit wurde $\gamma = 0,25$ angewendet.

LRP- z^β -Regel: Am Eingang des neuronalen Netzes wird in dieser Arbeit ein Bild übergeben. Die Pixel $\{x_p\}$ (hier in der allgemeinen Form $\{a_i\}$) dürfen nur bestimmte Werte annehmen. Die Berücksichtigung einer solchen Einschränkung bei der Suche nach der Entwicklungsstelle führt zur LRP- z^β -Regel. Die Regel für die Rückpropagation der Relevanz lautet dann [Mon+17]:

$$R_i = \sum_j \frac{z_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-}{\sum_i z_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-} \cdot R_j \text{ mit } z_{ij} = a_i w_{ij}, \quad (\text{A.4})$$

mit l_i , dem niedrigsten und h_i , dem höchsten Wert des Kanals des Bildes. Diese Regel wird nur verwendet, um die Relevanzwerte der Neuronen der ersten verdeckten Schicht zurück in den Eingangsraum zu propagieren, also für die Berechnung von $\{R_p\} = \{R_i\}^{(l=1)}$.

A.4.1 Kombination der Regeln

Wie von Montavon et al. [Mon+19] vorgeschlagen, werden die obigen Regeln kombiniert, um ihre jeweiligen Vor- und Nachteile auszugleichen. Die Architektur der visualisierten CNN-Klassifikatoren wurde in Abschnitt 5.1.2 erläutert. Die LRP-0-Regel wird für die vollständig verbundene Schicht, die die 256 Neuronen nach dem GAP mit den Ausgängen verknüpft, verwendet. In den Blöcken 3 und 4 wird die LRP- ϵ -Regel eingesetzt. In den Blöcken 1 und 2 kommt die LRP- γ -Regel zur Anwendung. Ergänzt werden diese drei Regeln durch die LRP- z^β -Regel, die für die Rückpropagation von der ersten Faltungsschicht auf den Eingangsraum verwendet wird, da man in diesem Fall den Wertebereich kennt.

A.5 Analyse weiterer Einflussgrößen auf die Grathöhe

In diesem Abschnitt wird untersucht und diskutiert, ob es einen systematischen Zusammenhang zwischen der Kante mit dem höchsten Grat und einer der folgenden Einflussgrößen gibt.

Schnitttrichtung Kante 1 weist bei 30 % aller Teile die maximale Grathöhe auf. Kante 4 bei 27 %, Kante 3 bei 24 % und Kante 2 bei 19 % der Teile. Es scheint also ein schwacher Zusammenhang zwischen der Schnitttrichtung und der Grathöhe zu bestehen. Ein Einfluss, der sich mit der Richtung überlagert, ist die Reihenfolge der Kanten. Der Strahl wird vor Kante 1 eingeschaltet, bleibt dann an und wird nach Kante 4 ausgeschaltet, damit der Schneidkopf zum nächsten Teil bewegt werden kann. Um beide Einflüsse zu trennen, müsste man die Kanten in einer anderen Reihenfolge schneiden.

Walzrichtung Bei Betrachtung der Walzrichtung werden die Kanten 1 und 3 bzw. 2 und 4 zusammengefasst. 1 und 3 wurden quer zur Walzrichtung des Blechs geschnitten, 2 und 4 parallel dazu. Die Differenz ist hier mit 54 % (maximale Grathöhe auf Kante 1 oder 3) und 46 % (maximale Grathöhe auf Kante 2 oder 4) sehr gering. Ein Einfluss, der sich mit der Walzrichtung überlagert, ist die Position der Stege. Diese sind in der Maschine orthogonal zur Walzrichtung angebracht und verändern (möglicherweise) die Gasdynamik. Um beide Einflüsse zu trennen, könnte man kleinere Blechtafeln verwenden und diese um 90° gedreht auf die Stege auflegen, sodass deren Walzrichtung entlang der Stege verläuft.

Blechtafel Die 899 Teile wurden aus insgesamt vier Blechen ausgeschnitten. Für die Teile 1 bis 250 wurde Blech 1 benutzt, die Teile 251 bis 550 wurden auf Blech 2 geschachtelt, für die Teile 551 bis 850 wurde Blech 3 und für die Teile 851 bis 1126 Blech 4 verwendet. Der Zusammenhang zwischen dem verwendeten Blech und der Kante mit der maximalen Grathöhe ist in Abbildung A.3(a) gezeigt. Kante 1 weist auf drei der vier Blechtafeln am häufigsten die größte Grathöhe auf, Kante 2 am seltensten. Nur Blech 4 passt nicht in diese Systematik, dort ist die Grathöhe von Kante 4 am

häufigsten die größte. Nichtsdestotrotz scheint das Blech einen eher geringen Einfluss zu haben. Da die Materialeigenschaften des Werkstoffs nicht im Detail bekannt sind, kann man von der Kenntnis der Zusammenhänge in den Blechen 1, 2, 3 und 4 auch nicht auf das Verhalten in einem weiteren Blech schließen. Eine mit dem Blech korrelierende Größe ist der Zeitpunkt des Schneidens, da die Bleche nacheinander geschnitten wurden.

Position auf dem Blech (Zeilen, Spalten) Auf eine Blechtafel der Länge 300 cm und der Breite 150 cm passen 25 Reihen bzw. 12 Spalten mit Teilen unter- bzw. nebeneinander (siehe Abbildung 4.2(b)). Es kommt teilweise zu deutlichen Unterschieden zwischen den Reihen, was die Kante mit der maximalen Grathöhe anbelangt. Während beispielsweise in Reihe 21 44 % der Teile auf Kante 4 die maximale Grathöhe aufweisen, sind es in Reihe 17 nur 11 %. Es ist jedoch kein systematischer Zusammenhang zwischen Reihe und Grathöhe erkennbar. Der Zusammenhang mit der Spalte ist in Abbildung A.3(b) gezeigt. Auffällig sind vor allem die Spalten 2 und 11, also die jeweils zweiten Spalten von außen. Während in Spalte 2 nur bei 11 % aller Teile der höchste Grat auf Kante 1 anzutreffen ist, sind es bei Spalte 11 40 %. Um eine fundiertere Aussage treffen zu können, müsste man die Position jedes einzelnen Teils sehr viel genauer erfassen, da diese davon abhängt, wie genau die Blechtafel auf die Palette gelegt wurde. Zusätzlich müsste man die Position der Stege und Stegspitzen vermessen.

Zeitpunkt des Schneidens, Maschinenlaufzeit Es konnte kein systematischer Zusammenhang zwischen der Kante mit dem größten Grat und dem Zeitpunkt des Schneidens bzw. der Maschinenlaufzeit festgestellt werden. Es ist also nicht davon auszugehen, dass die verschiedenen hohen Grate durch langsame thermische Einflüsse, wie die Erwärmung des Schneidkopfs durch Streustrahlung, entstehen.

Prozessparameter Die Prozessparameter beeinflussen die Grathöhe nicht nur direkt, sondern sie scheinen auch einen Einfluss darauf zu haben, an welcher Kante der Grat besonders ausgeprägt ist. In den Abbildungen A.3(c) und A.3(d) sind die Zusammenhänge mit dem Vorschub und der Fokusposition gezeigt. Beim Vorschub ist die Grathöhe für den LTT-Wert von 21 m/min besonders oft auf Kante 1 maximal. Große Grathöhen auf

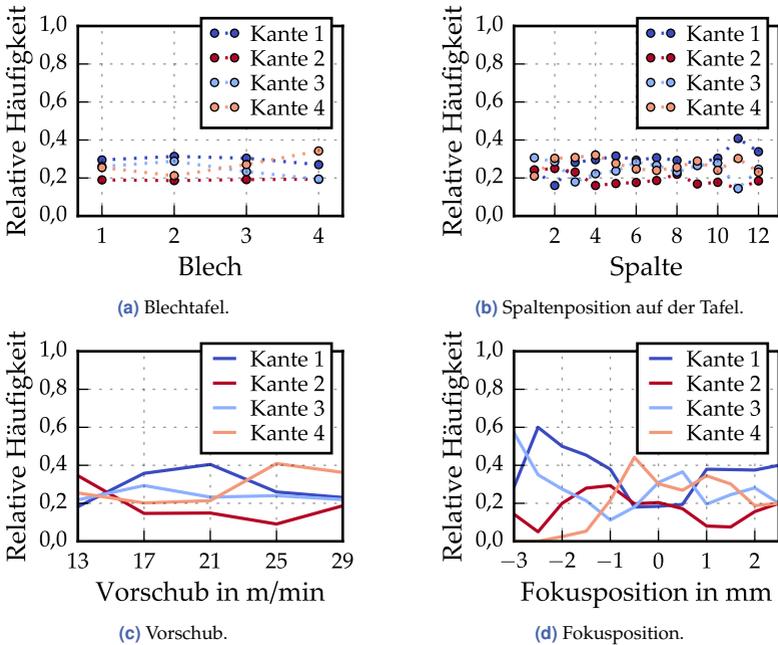


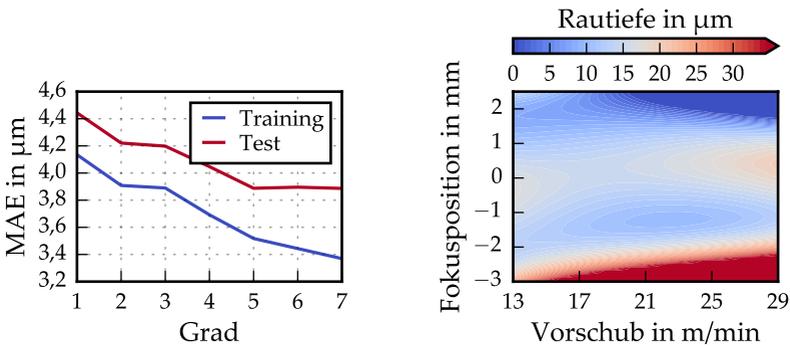
Abbildung A.3 Zusammenhang zwischen der Kante mit der maximalen Grathöhe und verschiedenen Einflussgrößen.

Kante 2 nehmen mit steigendem Vorschub eher ab, auf Kante 4 dagegen eher zu. Die maximale Grathöhe auf Kante 3 ist mit Anteilen von 21 % bis 29 % relativ stabil. Bei der Fokuslage sollte man die Extremwerte von $-3,0$ und $2,5$ mm aufgrund der geringen Datengrundlage außer Acht lassen. Erkennbare Tendenzen sind ein überdurchschnittlich niedriger Anteil von Kante 1 im Bereich des LTT-Werts von $-0,5$ mm. Kante 4 verhält sich gegenläufig. Die Gründe für dieses Verhalten sind jedoch unklar.

Des Weiteren wurde nach einem Zusammenhang zwischen den obenstehenden Einflüssen und der Grathöhe gesucht. Eindeutige Zusammenhänge konnten nicht identifiziert werden.

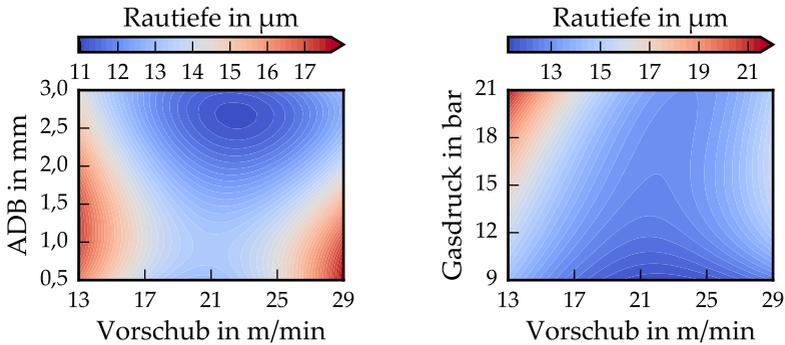
A.6 Modellierung der Rautiefe

Die folgenden Abbildungen A.4, A.5 und A.6 dienen als Ergänzung der in Abschnitt 6.4.3.1 präsentierten Ergebnisse. Sie zeigen die Zusammenhänge zwischen den Prozessparametern und der Rautiefe auf den Messlinien 03, 15 und 27.



(a) Mittlerer absoluter Fehler für verschiedene polynomiale Erweiterungen.

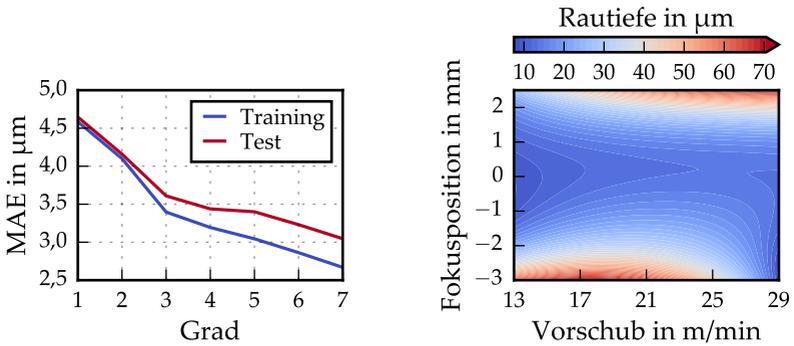
(b) ADB 1,5 mm, Gasdruck 15 bar.



(c) Fokusposition $-0,5$ mm, Gasdruck 15 bar.

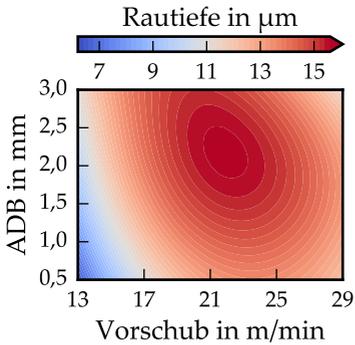
(d) ADB 1,5 mm, Fokusposition $-0,5$ mm.

Abbildung A.4 Güte verschiedener Modelle für die Prädiktion der Rautiefe auf Messlinie 03: (a). Mit dem Modell 4. Grades berechnete Rautiefe: (b), (c), (d).

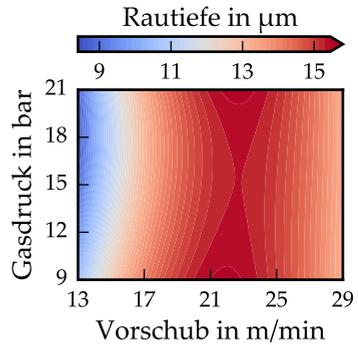


(a) Mittlerer absoluter Fehler für verschiedene polynomiale Erweiterungen.

(b) ADB 1,5 mm, Gasdruck 15 bar.

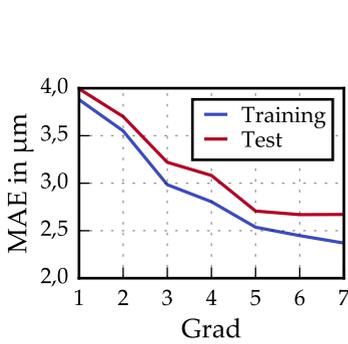


(c) Fokusposition $-0,5$ mm, Gasdruck 15 bar.

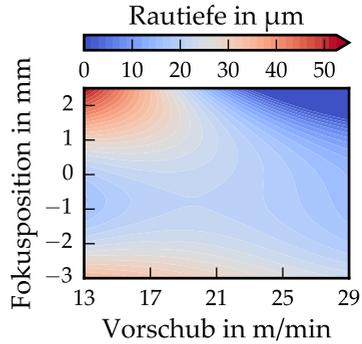


(d) ADB 1,5 mm, Fokusposition $-0,5$ mm.

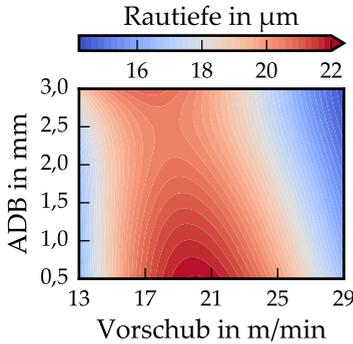
Abbildung A.5 Güte verschiedener Modelle für die Prädiktion der Rautiefe auf Messlinie 15: (a). Mit dem Modell 4. Grades berechnete Rautiefe: (b), (c), (d).



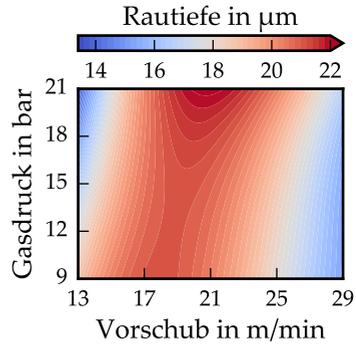
(a) Mittlerer absoluter Fehler für verschiedene polynomiale Erweiterungen.



(b) ADB 1,5 mm, Gasdruck 15 bar.



(c) Fokusposition -0,5 mm, Gasdruck 15 bar.



(d) ADB 1,5 mm, Fokusposition -0,5 mm.

Abbildung A.6 Güte verschiedener Modelle für die Prädiktion der Rautiefe auf Messlinie 27: (a). Mit dem Modell 4. Grades berechnete Rautiefe: (b), (c), (d).

Literaturverzeichnis

- [Aba+16] **Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. und Zheng, X.** *TensorFlow: Large-scale Machine Learning on Heterogeneous Distributed Systems*. In: *arXiv Preprint* (2016). arXiv: 1603.04467.
- [Agg17] **Aggarwal, C. C.** *Outlier Analysis*. 2. Aufl. Cham, Schweiz: Springer International Publishing, 2017.
- [Alb+19] **Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Samek, W., Müller, K.-R., Dähne, S. und Kindermans, P.-J.** *iNNvestigate Neural Networks!* In: *Journal of Machine Learning Research* 20.93 (2019), S. 1–8.
- [Ant+12] **Antunes, P., Herskovic, V., Ochoa, S. F. und Pino, J. A.** *Structuring Dimensions for Collaborative Systems Evaluation*. In: *ACM Computing Surveys* 44 (2012), S. 1–28.
- [Arr+16] **Arras, L., Horn, F., Montavon, G., Müller, K.-R. und Samek, W.** *Explaining Predictions of Non-Linear Classifiers in NLP*. In: *1st Workshop on Representation Learning for NLP*. Association for Computational Linguistics. Berlin, 2016, S. 1–7.
- [BK15] **Ba, J. und Kingma, D.** *Adam: A Method for Stochastic Optimization*. In: *3rd International Conference for Learning Representations*. San Diego, CA, USA, 2015.

- [Bac+15] **Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R. und Samek, W.** *On Pixel-wise Explanations for Non-linear Classifier Decisions by Layer-wise Relevance Propagation*. In: *PLoS ONE* 10.7 (2015), e0130140.
- [Bal+17] **Balduzzi, D., Frean, M., Leary, L., Lewis, J. P., Ma, K. W.-D. und McWilliams, B.** *The Shattered Gradients Problem: If ResNets are the Answer, then What is the Question?* In: *34th International Conference on Machine Learning*. JMLR.org. Sydney, NSW, Australien, 2017, S. 332–341.
- [Bau+17] **Bau, D., Zhou, B., Khosla, A., Oliva, A. und Torralba, A.** *Network Dissection: Quantifying Interpretability of Deep Visual Representations*. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA, 2017, S. 3319–3327.
- [BPF16] **Beyerer, J., Puente León, F. und Frese, C.** *Automatische Sichtprüfung*. 2. Aufl. Berlin, Heidelberg: Springer Vieweg, 2016.
- [Bis06] **Bishop, C. M.** *Pattern Recognition and Machine Learning*. 1. Aufl. New York, NY, USA: Springer, 2006.
- [Böh+19] **Böhle, M., Eitel, F., Weygandt, M. und Ritter, K.** *Layer-wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-based Alzheimer’s Disease Classification*. In: *Frontiers in Aging Neuroscience* 11.194 (2019), S. 194.
- [BMB18] **Borkmann, M., Mahrle, A. und Beyer, E.** *Study of Correlation between Edge Roughness and Gas Flow Characteristics in Laser Beam Fusion Cutting*. In: *10th CIRP Conference on Photonic Technologies*. Fürth, 2018, S. 421–424.
- [Bra00] **Bradski, G.** *The OpenCV Library*. 2000. URL: <https://www.drdoobs.com/open-source/the-opencv-library/184404319> (besucht am 29. 12. 2020).
- [BÖ15] **Brecher, C. und Özdemir, D.** *Introduction*. In: *Advances in Production Technology*. Hrsg. von **Brecher, C.** Cham, Schweiz: Springer International Publishing, 2015, S. 1–8.
- [BG11] **Bühlmann, P. und Geer, S. van de.** *Statistics for High-dimensional Data*. 1. Aufl. Berlin, Heidelberg: Springer, 2011.

- [Can86] **Canny, J.** *A Computational Approach to Edge Detection*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8.6 (1986), S. 679–698.
- [CBB20] **Chaki, S., Bose, D. und Bathe, R. N.** *Multi-objective Optimization of Pulsed Nd: YAG Laser Cutting Process Using Entropy-based ANN-PSO Model*. In: *Lasers in Manufacturing and Materials Processing* 7.1 (2020), S. 88–110.
- [CM87] **Chan, C. L. und Mazumder, J.** *One-dimensional Steady-state Model for Damage by Vaporization and Liquid Expulsion due to Laser-material Interaction*. In: *Journal of Applied Physics* 62.11 (1987), S. 4579–4586.
- [Cho15] **Chollet, F.** *Keras*. 2015. URL: <https://keras.io> (besucht am 26.02.2020).
- [CA77] **Cline, H. E. und Anthony, T. R.** *Heat Treating and Melting Material with a Scanning Laser or Electron Beam*. In: *Journal of Applied Physics* 48.9 (1977), S. 3895–3900.
- [Cro16] **CrowdFlower (heute FigureEight)**. *Data Science Report*. Towards Data Science. 2016. URL: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf (besucht am 05.12.2020).
- [Den+09] **Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. und Fei-Fei, L.** *ImageNet: A Large-scale Hierarchical Image Database*. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA, 2009, S. 248–255.
- [Deu98a] **Deutsches Institut für Normung e. V.** *DIN EN ISO 3274: Geometrische Produktspezifikationen (GPS) - Oberflächenbeschaffenheit: Tastschnittverfahren - Nenneigenschaften von Tastschnittgeräten*. 1998.
- [Deu98b] **Deutsches Institut für Normung e. V.** *DIN EN ISO 4288: Geometrische Produktspezifikation (GPS) - Oberflächenbeschaffenheit: Tastschnittverfahren - Regeln und Verfahren für die Beurteilung der Oberflächenbeschaffenheit*. 1998.

- [Deu10a] **Deutsches Institut für Normung e. V.** *DIN EN 10088-4: Nichtrostende Stähle - Teil 4: Technische Lieferbedingungen für Blech und Band aus korrosionsbeständigen Stählen für das Bauwesen.* 2010.
- [Deu10b] **Deutsches Institut für Normung e. V.** *DIN EN ISO 9445-2: Kontinuierlich kaltgewalzter nichtrostender Stahl – Grenzabmaße und Formtoleranzen – Teil 2: Kaltbreitband und Blech.* 2010.
- [Deu14] **Deutsches Institut für Normung e. V.** *DIN EN ISO 10088-2: Nichtrostende Stähle - Teil 2: Technische Lieferbedingungen für Blech und Band aus korrosionsbeständigen Stählen für allgemeine Verwendung.* 2014.
- [Deu15] **Deutsches Institut für Normung e. V.** *DIN EN 60825-1: Sicherheit von Lasereinrichtungen - Teil 1: Klassifizierung von Anlagen und Anforderungen.* 2015.
- [Deu16] **Deutsches Institut für Normung e. V.** *DIN EN 10028-7: Flacherzeugnisse aus Druckbehälterstählen - Teil 7: Nichtrostende Stähle.* 2016.
- [Deu17] **Deutsches Institut für Normung e. V.** *DIN EN ISO 9013: Thermisches Schneiden – Einteilung thermischer Schnitte – Geometrische Produktspezifikation und Qualität.* 2017.
- [DHS00] **Duda, R. O., Hart, P. E. und Stork, D. G.** *Pattern Classification.* 2000. Aufl. New York, NY, USA: Wiley, 2000.
- [DV18] **Dumoulin, V. und Visin, F.** *A Guide to Convolution Arithmetic for Deep Learning.* In: *arXiv Preprint* (2018). arXiv: 1603.07285.
- [Ein17] **Einstein, A.** *The Quantum Theory of Radiation.* In: *Physikalische Zeitschrift* 18 (1917), S. 121–135.
- [Est+17] **Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. und Thrun, S.** *Dermatologist-level Classification of Skin Cancer with Deep Neural Networks.* In: *Nature* 542.7639 (2017), S. 115–118.
- [Eur17] **Eurostat.** *Manufacturing Statistics - NACE Rev. 2.* 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Manufacturing_statistics_-_NACE_Rev._2 (besucht am 21.07.2020).

- [Eve+10] **Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. und Zisserman, A.** *The Pascal Visual Object Classes Challenge*. In: *International Journal of Computer Vision* 88.2 (2010), S. 303–338.
- [Fri14] **Fricke, K.** *Digitaltechnik*. 7. Aufl. Wiesbaden: Springer Vieweg, 2014.
- [GSP20] **Gonçalves, D. A., Stemmer, M. R. und Pereira, M.** *A Convolutional Neural Network Approach on Bead Geometry Estimation for a Laser Cladding System*. In: *The International Journal of Advanced Manufacturing Technology* 106.5 (2020), S. 1811–1821.
- [Gre+20] **Grezmak, J., Zhang, J., Wang, P., Loparo, K. A. und Gao, R. X.** *Interpretable Convolutional Neural Network Through Layer-wise Relevance Propagation for Machine Fault Diagnosis*. In: *IEEE Sensors Journal* 20.6 (2020), S. 3172–3181.
- [HJ17] **Habibi Aghdam, H. und Jahani Heravi, E.** *Guide to Convolutional Neural Networks*. 1. Aufl. Cham, Schweiz: Springer, 2017.
- [Har+20] **Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. und Oliphant, T. E.** *Array Programming with NumPy*. In: *Nature* 585.7825 (2020), S. 357–362.
- [HTF09] **Hastie, T., Tibshirani, R. und Friedman, J.** *The Elements of Statistical Learning*. 2. Aufl. New York, NY, USA: Springer, 2009.
- [He+15] **He, K., Zhang, X., Ren, S. und Sun, J.** *Delving Deep into Rectifiers: Surpassing Human-level Performance on ImageNet Classification*. In: *IEEE International Conference on Computer Vision*. Santiago, Chile, 2015, S. 1026–1034.
- [He+16] **He, K., Zhang, X., Ren, S. und Sun, J.** *Deep Residual Learning for Image Recognition*. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 2016, S. 770–778.

- [HSR14] **Hesse, T., Schindhelm, D. und Regaard, B.** *Method and Device for Detecting a Cut Section*. Europ. Pat. 2805791A3. **TRUMPF Werkzeugmaschinen GmbH + Co. KG**. 2014.
- [Hig12] **High, R.** *The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works*. 2012. URL: <http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf> (besucht am 29. 12. 2020).
- [Hin06] **Hinton, G. E.** *Reducing the Dimensionality of Data with Neural Networks*. In: *Science* 313.5786 (2006), S. 504–507.
- [Hin+12] **Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. und R. Salakhutdinov, R.** *Improving Neural Networks by Preventing Co-adaptation of Feature Detectors*. In: *arXiv Preprint* (2012). arXiv: 1207.0580v1.
- [HG09] **Hügel, H. und Graf, T.** *Laser in der Fertigung*. 2. Aufl. Wiesbaden: Vieweg + Teubner, 2009.
- [Hun07] **Hunter, J. D.** *Matplotlib: A 2D Graphics Environment*. In: *Computing in Science & Engineering* 9.3 (2007), S. 90–95.
- [IS15] **Ioffe, S. und Szegedy, C.** *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. In: *32nd International Conference on Machine Learning*. JMLR.org. Lille, Frankreich, 2015, S. 448–456.
- [Iva+94] **Ivarson, A., Powell, J., Kamalu, J. und Magnusson, C.** *The Oxidation Dynamics of Laser Cutting of Mild Steel and the Generation of Striations on the Cut Edge*. In: *Journal of Materials Processing Technology* 40.3 (1994), S. 359–374.
- [Jam+13] **James, G., Witten, D., Hastie, T. und Tibshirani, R.** *An Introduction to Statistical Learning*. 1. Aufl. New York, NY, USA: Springer, 2013.
- [KRL97] **Kar, A., Rothenflue, J. A. und Latham, W. P.** *Scaling Laws for Thick-section Cutting with a Chemical Oxygen–Iodine Laser*. In: *Journal of Laser Applications* 9.6 (1997), S. 279–286.

- [KYP17] **Khumaidi, A., Yuniarno, E. M. und Purnomo, M. H.** *Welding Defect Classification Based on Convolution Neural Network and Gaussian Kernel*. In: *2017 International Seminar on Intelligent Technology and Its Applications*. Surabaya, Indonesien, 2017, S. 261–265.
- [KC19] **Kim, S. und Choi, J.-H.** *Convolutional Neural Network for Gear Fault Diagnosis Based on Signal Segmentation Approach*. In: *Structural Health Monitoring* 18.5 (2019), S. 1401–1415.
- [Kin+18] **Kincade, K., Nogee, A., Overton, G., Belforte, D. und Holton, C.** *Annual Laser Market Review & Forecast: Lasers Enabling Lasers*. 2018. URL: <https://www.laserfocusworld.com/articles/print/volume-54/issue-01/features/annual-laser-market-review-forecast-lasers-enabling-lasers.html> (besucht am 05.04.2019).
- [KL17] **Koh, P. W. und Liang, P.** *Understanding Black-box Predictions via Influence Functions*. In: *34th International Conference on Machine Learning*. JMLR.org. Sydney, NSW, Australien, 2017, S. 1885–1894.
- [KK01] **Kolen, J. F. und Kremer, S. C.** *A Field Guide to Dynamical Recurrent Networks*. New York, NY, USA: Wiley-IEEE Press, 2001.
- [KSH17] **Krizhevsky, A., Sutskever, I. und Hinton, G. E.** *ImageNet Classification with Deep Convolutional Neural Networks*. In: *Communications of the ACM* 60.6 (2017), S. 84–90.
- [Kub15] **Kubat, M.** *An Introduction to Machine Learning*. 1. Aufl. Cham, Schweiz: Springer International Publishing, 2015.
- [KJ13] **Kuhn, M. und Johnson, K.** *Applied Predictive Modeling*. 1. Aufl. New York, NY, USA: Springer, 2013.
- [Kus18] **Kusiak, A.** *Smart Manufacturing*. In: *International Journal of Production Research* 56.1 (2018), S. 508–517.
- [Lap19] **Lapuschkina, S.** *Opening the Machine Learning Black Box with Layer-wise Relevance Propagation*. Doktorarbeit. Technische Universität Berlin, 2019.

- [Lap+19] **Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W. und Müller, K.-R.** *Unmasking Clever Hans Predictors and Assessing what Machines Really Learn*. In: *Nature Communications* 10.1, 1096 (2019).
- [LeC+98] **LeCun, Y., Bottou, L., Bengio, Y. und Haffner, P.** *Gradient-based Learning Applied to Document Recognition*. In: *Proceedings of the IEEE* 86.11 (1998), S. 2278–2324.
- [LBH15] **LeCun, Y., Bengio, Y. und Hinton, G.** *Deep Learning*. In: *Nature* 521.7553 (2015), S. 436–444.
- [LKF10] **LeCun, Y., Kavukcuoglu, K. und Farabet, C.** *Convolutional Networks and Applications in Vision*. In: *2010 IEEE International Symposium on Circuits and Systems*. Paris, Frankreich, 2010, S. 253–256.
- [LeC+12] **LeCun, Y., Bottou, L., Orr, G. B. und Müller, K.-R.** *Efficient BackProp*. In: *Neural Networks: Tricks of the Trade*. Hrsg. von **Montavon, G., Orr, G. B. und Müller, K.-R.** Bd. 7700. Berlin, Heidelberg: Springer, 2012, S. 9–48.
- [LCK17] **Lee, K. B., Cheon, S. und Kim, C. O.** *A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes*. In: *IEEE Transactions on Semiconductor Manufacturing* 30.2 (2017), S. 135–142.
- [Lei08] **Leibinger, P.** *Über Kunden, Konkurrenten und Karrieren*. In: *Laser Technik Journal* 5.5 (2008), S. 9–12.
- [Lin+19] **Lin, H., Li, B., Wang, X., Shu, Y. und Niu, S.** *Automated Defect Inspection of LED Chip Using Deep Convolutional Neural Network*. In: *Journal of Intelligent Manufacturing* 30.6 (2019), S. 2525–2534.
- [LCY14] **Lin, M., Chen, Q. und Yan, S.** *Network In Network*. In: *2nd International Conference on Learning Representations*. Banff, AB, Kanada, 2014.
- [LRZ06] *Einführung*. In: *Individualisierte Produkte — Komplexität beherrschen in Entwicklung und Produktion*. Hrsg. von **Lindemann, U., Reichwald, R. und Zäh, M. F.** Berlin, Heidelberg: Springer, 2006.

- [Liu+20] **Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R. und Cui, Y.** *Improved Kiwifruit Detection Using Pre-trained VGG16 with RGB and NIR Information Fusion*. In: *IEEE Access* 8 (2020), S. 2327–2336.
- [MR12a] **Madić, M. und Radovanović, M.** *An Artificial Intelligence Approach for the Prediction of Surface Roughness in CO₂ Laser Cutting*. In: *Journal of Engineering Science and Technology* 7.6 (2012), S. 679–689.
- [MR12b] **Madić, M. und Radovanović, M.** *Analysis of the Heat Affected Zone in CO₂ Laser Cutting of Stainless Steel*. In: *Thermal Science* 16.2 (2012), S. 363–373.
- [MR12c] **Madić, M. und Radovanović, M.** *Comparative Modeling of CO₂ Laser Cutting Using Multiple Regression Analysis and Artificial Neural Network*. In: *International Journal of Physical Sciences* 7.16 (2012), S. 2422–2430.
- [Mai60] **Maiman, T. H.** *Stimulated Optical Radiation in Ruby*. In: *Nature* 187.4736 (1960), S. 493–494.
- [Mal+19] **Malin, L., Jansen, A., Seyda, S. und Flake, R.** *Fachkräftesicherung in Deutschland – diese Potenziale gibt es noch*. 2019. URL: https://www.kofa.de/fileadmin/Dateiliste/Publikationen/Studien/Fachkraefteengaesse_2019_2.pdf (besucht am 09.01.2021).
- [MP43] **McCulloch, W. S. und Pitts, W.** *A Logical Calculus of the Ideas Immanent in Nervous Activity*. In: *The Bulletin of Mathematical Biophysics* 5.4 (1943), S. 115–133.
- [McK10] **McKinney, W.** *Data Structures for Statistical Computing in Python*. Austin, TX, USA, 2010.
- [Mer+15] **Mery, D., Riffo, V., Zscherpel, U., Mondragón, G., Lillo, I., Zuc-car, I., Lobel, H. und Carrasco, M.** *GD_Xray: The Database of X-ray Images for Nondestructive Testing*. In: *Journal of Nondestructive Evaluation* 34.4, 42 (2015).
- [Mit97] **Mitchell, T. M.** *Machine Learning*. 1. Aufl. New York, NY, USA: McGraw-Hill, 1997.

- [Moh94] **Mohr, U.** *Geschwindigkeitsbestimmende Strahleigenschaften und Einkoppelmechanismen beim CO₂-Laserschneiden von Metallen*. 1. Aufl. Wiesbaden: Vieweg + Teubner, 1994.
- [MSM18] **Montavon, G., Samek, W. und Müller, K.-R.** *Methods for Interpreting and Understanding Deep Neural Networks*. In: *Digital Signal Processing* 73 (2018), S. 1–15.
- [Mon+15] **Montavon, G., Lapuschkin, S., Binder, A., Samek, W. und Müller, K.-R.** *Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition*. In: *arXiv Preprint* (2015). arXiv: 1512.02479.
- [Mon+17] **Montavon, G., Lapuschkin, S., Binder, A., Samek, W. und Müller, K.-R.** *Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition*. In: *Pattern Recognition* 65 (2017), S. 211–222.
- [Mon+19] **Montavon, G., Binder, A., Lapuschkin, S., Samek, W. und Müller, K.-R.** *Layer-wise Relevance Propagation: An Overview*. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Hrsg. von **Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. und Müller, K.-R.** Bd. 11700. Cham, Schweiz: Springer International Publishing, 2019, S. 193–209.
- [Mor+95] **Morch, N., Kjems, U., Hansen, L., Svarer, C., Law, I., Lautrup, B., Strother, S. und Rehm, K.** *Visualization of Neural Networks Using Saliency Maps*. In: *International Conference on Neural Networks*. IEEE. Perth, WA, Australien, 1995, S. 2085–2090.
- [Mül20] **Müller, C.** *Weltmarktführerindex*. 2020. URL: https://www.weltmarktfuehrerindex.de/corp_profile/?pid=226 (besucht am 10. 07. 2020).
- [Mur12] **Murphy, K. P.** *Machine Learning: A Probabilistic Perspective*. 1. Aufl. Cambridge, MA, USA: MIT Press, 2012.
- [Ols11] **Olsen, F. O.** *Laser Cutting from CO₂ Laser to Disc or Fiber Laser – Possibilities and Challenges*. In: *30th International Congress on Laser Materials Processing, Laser Microprocessing and Nanomanufacturing*. Laser Institute of America. Orlando, FL, USA, 2011, S. 6–15.

- [Ott+11] **Otto, A., Koch, H., Leitz, K.-H. und Schmidt, M.** *Numerical Simulations – A Versatile Approach for Better Understanding Dynamics in Laser Material Processing*. In: *Physics Procedia* 12 (2011), S. 11–20.
- [Pan19] **Pant, A.** *Towards Data Science: Workflow of a Machine Learning Project*. 2019. URL: <https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94> (besucht am 05. 12. 2020).
- [Ped+11] **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. und Louppe, G.** *Scikit-Learn: Machine Learning in Python*. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830.
- [Pet16] **Petring, D.** *Virtual Laser Cutting Simulation for Real Parameter Optimization*. In: *84th Laser Materials Processing Conference*. Japan Laser Processing Society. Nagoya, Japan, 2016, S. 11–20.
- [PSW12] **Petring, D., Schneider, F. und Wolf, N.** *Some Answers to Frequently Asked Questions and Open Issues of Laser Beam Cutting*. In: *31st International Congress on Laser Materials Processing, Laser Microprocessing and Nanomanufacturing*. Laser Institute of America. Anaheim, CA, USA, 2012, S. 43–48.
- [Pin99] **Pinkus, A.** *Approximation Theory of the MLP Model in Neural Networks*. In: *Acta Numerica* 8 (1999), S. 143–195.
- [PSS10] **Poprawe, R., Schulz, W. und Schmitt, R.** *Hydrodynamics of Material Removal by Melt Expulsion: Perspectives of Laser Cutting and Drilling*. In: *Physics Procedia* 5 (2010), S. 1–18.
- [PW89] **Powell, J. und Wykes, C.** *A Comparison Between CO₂ Laser Cutting and Competitive Techniques*. In: *6th International Conference on Lasers in Manufacturing*. Birmingham, England, 1989, S. 135–153.
- [Al+17] **Al-Qizwini, M., Barjasteh, I., Al-Qassab, H. und Radha, H.** *Deep Learning Algorithm for Autonomous Driving Using GoogLeNet*. In: *2017 IEEE Intelligent Vehicles Symposium*. Los Angeles, CA, USA, 2017, S. 89–96.

- [RSG16] **Ribeiro, M. T., Singh, S. und Guestrin, C.** "Why Should I Trust You?": *Explaining the Predictions of Any Classifier*. In: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA, 2016, S. 1135–1144.
- [Ros46] **Rosenthal, D.** *The Theory of Moving Sources of Heat and its Application to Metal Treatments*. In: *Trans ASME* 48 (1946), S. 849–866.
- [RHW86] **Rumelhart, D. E., Hinton, G. E. und Williams, R. J.** *Learning Representations by Back-propagating Errors*. In: *Nature* 323.6088 (1986), S. 533–536.
- [RND10] **Russell, S. J., Norvig, P. und Davis, E.** *Artificial Intelligence: A Modern Approach*. 3. Aufl. Upper Saddle River, NJ, USA: Prentice Hall, 2010.
- [SM19] **Samek, W. und Müller, K.-R.** *Towards Explainable Artificial Intelligence*. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Hrsg. von **Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. und Müller, K.-R.** Bd. 11700. Cham, Schweiz: Springer International Publishing, 2019, S. 5–22.
- [Sam59] **Samuel, A. L.** *Some Studies in Machine Learning Using the Game of Checkers*. In: *IBM Journal of Research and Development* 3 (1959), S. 210–229.
- [San+19] **Santurkar, S., Tsipras, D., Ilyas, A. und Madry, A.** *How Does Batch Normalization Help Optimization?* In: *32nd Conference on Neural Information Processing Systems*. Montréal, QC, Kanada, 2019, S. 2483–2493.
- [SZL16] **Sateesh Babu, G., Zhao, P. und Li, X.-L.** *Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life*. In: *Database Systems for Advanced Applications*. Hrsg. von **Navathe, S. B., Wu, W., Shekhar, S., Du, X., Wang, X. S. und Xiong, H.** Bd. 9642. Cham, Schweiz: Springer International Publishing, 2016, S. 214–228.

- [SMB10] **Scherer, D., Müller, A. und Behnke, S.** *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition*. In: *Artificial Neural Networks – ICANN 2010*. Hrsg. von **Diamantaras, K., Duch, W. und Iliadis, L. S.** Bd. 6354. Berlin, Heidelberg: Springer, 2010, S. 92–101.
- [SGK17] **Shrikumar, A., Greenside, P. und Kundaje, A.** *Learning Important Features Through Propagating Activation Differences*. In: *34th International Conference on Machine Learning*. JMLR.org. Sydney, NSW, Australien, 2017, S. 3145–3153.
- [Sim07] **Simon, H.** *Hidden Champions des 21. Jahrhunderts*. 1. Aufl. Frankfurt am Main: Campus, 2007.
- [SVZ14] **Simonyan, K., Vedaldi, A. und Zisserman, A.** *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. In: *2nd International Conference on Learning Representations*. Banff, AB, Kanada, 2014.
- [SZ15] **Simonyan, K. und Zisserman, A.** *Very Deep Convolutional Networks for Large-scale Image Recognition*. In: *3rd International Conference on Learning Representations*. San Diego, CA, USA, 2015.
- [Siz+19] **Sizyakin, R., Voronin, V. V., Gapon, N., Zelensky, A. und Pižurica, A.** *Automatic Detection of Welding Defects Using the Convolutional Neural Network*. In: *Automated Visual Inspection and Machine Vision*. International Society for Optics und Photonics. München, 2019, 93–101.
- [SSA08] **Solomatine, D., See, L. und Abrahart, R.** *Data-driven Modeling: Concepts, Approaches and Experiences*. In: *Practical Hydroinformatics*. Hrsg. von **Abrahart, R. J., See, L. M. und Solomatine, D. P.** Bd. 68. Berlin, Heidelberg: Springer, 2008, S. 17–30.
- [Spr+15] **Springenberg, J. T., Dosovitskiy, A., Brox, T. und Riedmiller, M.** *Striving for Simplicity: The All Convolutional Net*. In: *3rd International Conference on Learning Representations*. San Diego, CA, USA, 2015.
- [SGS15] **Srivastava, R. K., Greff, K. und Schmidhuber, J.** *Training Very Deep Networks*. In: *29th Conference on Neural Information Processing Systems*. Montréal, QC, Kanada, 2015, S. 2377–2385.

- [SJ19] **Stahl, J. und Jauch, C.** *Quick Roughness Evaluation of Cut Edges Using a Convolutional Neural Network*. In: *14th International Conference on Quality Control by Artificial Vision*. International Society for Optics und Photonics. München, 2019, S. 175–181.
- [SM10] **Steen, W. M. und Mazumder, J.** *Laser Material Processing*. 4. Aufl. London, England: Springer, 2010.
- [Sti19] **Stix, L.** *Ermittlung von Anforderungen an die Substitution von Bearbeitungsschritten in der Fertigung lasergeschnittener Blechteile*. Bachelorarbeit. Hochschule Furtwangen, 2019.
- [SH67] **Sullivan, A. und Houldcroft, P.** *Gas-jet Laser Cutting*. In: *British Welding Journal* 14 (1967), S. 443–445.
- [Sun+18] **Sun, W., Yao, B., Chen, B., He, Y., Cao, X., Zhou, T. und Liu, H.** *Noncontact Surface Roughness Estimation Using 2D Complex Wavelet Enhanced ResNet for Intelligent Evaluation of Milled Metal Surface Quality*. In: *Applied Sciences* 8.3, 381 (2018).
- [Sve+12] **Svelto, O., Longhi, S., Valle, G. D., Huber, G., Kück, S., Pollnau, M., Hillmer, H., Kusserow, T., Engelbrecht, R., Rohlfing, F., Kaiser, J., Malz, R., Marowsky, G., Mann, K., Simon, P., Rhodes, C. K., Duarte, F. J., Borsutzky, A., L’huillier, J. A., Sigrist, M. W., Wächter, H., Saldin, E., Schneidmiller, E., Yurkov, M., Sauerbrey, R., Hein, J., Gianella, M., Helmcke, J., Midorikawa, K., Riehle, F., Steinberg, S. und Brand, H.** *Lasers and Coherent Light Sources*. In: *Springer Handbook of Lasers and Optics*. Hrsg. von **Träger, F.** Berlin, Heidelberg: Springer, 2012, S. 641–1046.
- [Tai+14] **Taigman, Y., Yang, M., Ranzato, M. und Wolf, L.** *DeepFace: Closing the Gap to Human-level Performance in Face Verification*. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA, 2014, S. 1701–1708.
- [Ter+16] **Tercan, H., Khawli, T. A., Eppelt, U., Büscher, C., Meisen, T. und Jeschke, S.** *Use of Classification Techniques to Design Laser Cutting Processes*. In: *The 6th International Conference on Changeable, Agile, Reconfigurable and Virtual Production*. CIRP. Bath, England, 2016, S. 292–297.

- [Ter+17] **Tercan, H., Khawli, T. A., Eppelt, U., Büscher, C., Meisen, T. und Jeschke, S.** *Improving the Laser Cutting Process Design by Machine Learning Techniques*. In: *Production Engineering* 11.2 (2017), S. 195–203.
- [Tho+14] **Thombansen, U., Hermanns, T., Molitor, T., Pereira, M. und Schulz, W.** *Measurement of Cut Front Properties in Laser Cutting*. In: *Physics Procedia* 56 (2014), S. 885–891.
- [TLC08] **Tsai, M.-J., Li, C.-H. und Chen, C.-C.** *Optimal Laser-cutting Parameters for QFN Packages by Utilizing Artificial Neural Networks and Genetic Algorithm*. In: *Journal of Materials Processing Technology* 208.1 (2008), S. 270–283.
- [Vin+10] **Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. und Manzagol, P.-A.** *Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion*. In: *The Journal of Machine Learning Research* 11 (2010), S. 3371–3408.
- [Vis+21] **Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A. und Plöd, M.** *An Overview of the End-to-end Machine Learning Workflow*. 2021. URL: <https://ml-ops.org/content/end-to-end-ml-workflow> (besucht am 12.02.2021).
- [WK10] **Wandera, C. und Kujanpää, V.** *Characterization of the Melt Removal Rate in Laser Cutting of Thick-section Stainless Steel*. In: *Journal of Laser Applications* 22.2 (2010), S. 62–70.
- [WK11] **Wandera, C. und Kujanpää, V.** *Optimization of Parameters for Fibre Laser Cutting of a 10 mm Stainless Steel Plate*. In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 225.5 (2011), S. 641–649.
- [WYH12] **Wang, P., Youn, B. D. und Hu, C.** *A generic probabilistic framework for structural health prognostics and uncertainty management*. In: *Mechanical Systems and Signal Processing* 28 (2012), S. 622–637.
- [Wue+16] **Wuest, T., Weimer, D., Irgens, C. und Thoben, K.-D.** *Machine Learning in Manufacturing: Advantages, Challenges, and Applications*. In: *Production & Manufacturing Research* 4.1 (2016), S. 23–45.

- [YA06] **Yilbas, B. S. und Aleem, B. J. A.** *Dross Formation During Laser Cutting Process*. In: *Journal of Physics D: Applied Physics* 39.7 (2006), S. 1451–1461.
- [Yil17] **Yilbas, B. S.** *The Laser Cutting Process*. Oxford, England: Elsevier, 2017.
- [ZF14] **Zeiler, M. D. und Fergus, R.** *Visualizing and Understanding Convolutional Networks*. In: *Computer Vision – ECCV 2014*. Hrsg. von **Fleet, D., Pajdla, T., Schiele, B. und Tuytelaars, T.** Bd. 8689. Cham, Schweiz: Springer International Publishing, 2014, S. 818–833.
- [ZWC19] **Zhang, Z., Wen, G. und Chen, S.** *Weld Image Deep Learning-based On-line Defects Detection Using Convolutional Neural Networks for Al Alloy in Robotic Arc Welding*. In: *Journal of Manufacturing Processes* 45 (2019), S. 208–216.
- [Zho+16] **Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. und Torralba, A.** *Learning Deep Features for Discriminative Localization*. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 2016, S. 2921–2929.

Eigene Veröffentlichungen

- [Blö+20] **Blömker, C., Kiefer, M., Mienhardt, U., Ottnad, J., Tatzel, L. und Weiß, K.** *Verfahren und Vorrichtung zur Bearbeitung eines Werkstücks*. Dt. Pat. 102018216873.3. **TRUMPF Werkzeugmaschinen GmbH + Co. KG**. Offenlegung 2020.
- [TKO20] **Tatzel, L., Kiefer, M. und Ottnad, J.** *Verfahren zum Bewerten einer Laserschnittkante, mobiles Endgerät und System*. Dt. Pat. 102019209088.5. **TRUMPF Werkzeugmaschinen GmbH + Co. KG**. Offenlegung 2020.
- [TO21a] **Tatzel, L. und Ottnad, J.** *Verfahren und Vorrichtung zum Ermitteln von Schneidparametern für eine Laserschneidmaschine*. Dt. Pat. 102019220478.3. **TRUMPF Werkzeugmaschinen GmbH + Co. KG**. Offenlegung 2021.

- [TO21b] **Tatzel, L. und Ottnad, J.** *Verfahren zur Bestimmung und Korrektur des Maschinenzustands einer Werkzeugmaschine und Diagnosesystem.* Dt. Pat. 102019220485.6. **TRUMPF Werkzeugmaschinen GmbH + Co. KG.** Offenlegung 2021.
- [TP19] **Tatzel, L. und Puente León, F.** *Prediction of Cutting Interruptions for Laser Cutting Using Logistic Regression.* In: *Lasers in Manufacturing.* Wissenschaftliche Gesellschaft Lasertechnik e.V. München, 2019.
- [TP20a] **Tatzel, L. und Puente León, F.** *Image-based Roughness Estimation of Laser Cut Edges with a Convolutional Neural Network.* In: *Procedia CIRP 94* (2020), S. 469–473.
- [TP20b] **Tatzel, L. und Puente León, F.** *Impact of the Thermally Induced Focus Shift on the Quality of a Laser Cutting Edge.* In: *Journal of Laser Applications 32.2*, 022022 (2020).
- [TP20c] **Tatzel, L. und Puente León, F.** *Models for Emulating the Human Quality Assessment of Laser Cut Edges.* In: *Procedia CIRP 99* (2020), S. 502–507.
- [Tat+21] **Tatzel, L., Hauaise, T., Al Tamimi, O. und Puente León, F.** *Image-based Modelling and Visualisation of the Relationship between Laser-Cut Edge and Process Parameters.* In: *Optics & Laser Technology 141* (2021), S. 107028.

Betreute studentische Arbeiten

- [Tan19] **Tang, Z.** *Objektivierung der Qualitätsbewertung bei 2D-Laserschnittkanten mit maschinellem Lernen.* Masterarbeit. Karlsruher Institut für Technologie, 2019.

Forschungsberichte aus der Industriellen Informationstechnik (ISSN 2190-6629)

- Band 1 Pérez Grassi, Ana
Variable illumination and invariant features for detecting and classifying varnish defects.
ISBN 978-3-86644-537-6
- Band 2 Christ, Konrad
Kalibrierung von Magnet-Injektoren für Benzin-Direkteinspritzsysteme mittels Körperschall.
ISBN 978-3-86644-718-9
- Band 3 Sandmair, Andreas
Konzepte zur Trennung von Sprachsignalen in unterbestimmten Szenarien.
ISBN 978-3-86644-744-8
- Band 4 Bauer, Michael
Vergleich von Mehrträger-Übertragungsverfahren und Entwurfskriterien für neuartige Powerline-Kommunikationssysteme zur Realisierung von Smart Grids.
ISBN 978-3-86644-779-0
- Band 5 Kruse, Marco
Mehrobjekt-Zustandsschätzung mit verteilten Sensorträgern am Beispiel der Umfeldwahrnehmung im Straßenverkehr.
ISBN 978-3-86644-982-4
- Band 6 Dudeck, Sven
Kamerabasierte In-situ-Überwachung gepulster Laserschweißprozesse.
ISBN 978-3-7315-0019-3
- Band 7 Liu, Wenqing
Emulation of Narrowband Powerline Data Transmission Channels and Evaluation of PLC Systems.
ISBN 978-3-7315-0071-1

- Band 8 Otto, Carola
Fusion of Data from Heterogeneous Sensors with Distributed Fields of View and Situation Evaluation for Advanced Driver Assistance Systems.
ISBN 978-3-7315-0073-5
- Band 9 Wang, Limeng
Image Analysis and Evaluation of Cylinder Bore Surfaces in Micrographs.
ISBN 978-3-7315-0239-5
- Band 10 Michelsburg, Matthias
Materialklassifikation in optischen Inspektionssystemen mithilfe hyperspektraler Daten.
ISBN 978-3-7315-0273-9
- Band 11 Pallauf, Johannes
Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern.
ISBN 978-3-7315-0529-7
- Band 12 Sigle, Martin
Robuste Schmalband-Powerline-Kommunikation für Niederspannungsverteilternetze.
ISBN 978-3-7315-0539-6
- Band 13 Opalko, Oliver
Powerline-Kommunikation für Batteriemangement-Systeme in Elektro- und Hybridfahrzeugen.
ISBN 978-3-7315-0647-8
- Band 14 Han, Bin
Characterization and Emulation of Low-Voltage Power Line Channels for Narrowband and Broadband Communication.
ISBN 978-3-7315-0654-6
- Band 15 Alonso, Damián Ezequiel
Wireless Data Transmission for the Battery Management System of Electric and Hybrid Vehicles.
ISBN 978-3-7315-0670-6

- Band 16 Hernández Mesa, Pilar
Design and analysis of a content-based image retrieval system.
ISBN 978-3-7315-0692-8
- Band 17 Suchanek, André
Energiemanagement-Strategien für batterieelektrische Fahrzeuge.
ISBN 978-3-7315-0773-4
- Band 18 Bauer, Sebastian
Hyperspectral Image Unmixing Incorporating Adjacency Information.
ISBN 978-3-7315-0788-8
- Band 19 Vater, Sebastian
Monokulare Blickrichtungsschätzung zur berührungslosen Mensch-Maschine-Interaktion.
ISBN 978-3-7315-0789-5
- Band 20 Back, Kristine
Erkennung menschlicher Aktivitäten durch Erfassung und Analyse von Bewegungstrajektorien.
ISBN 978-3-7315-0909-7
- Band 21 Nürnberg, Thomas
Entwurf von Computational-Imaging-Systemen am Beispiel der monokularen Tiefenschätzung.
ISBN 978-3-7315-0941-7
- Band 22 Kaiser, Cornelius
Adaptive Modulationsverfahren für die schmalbandige Powerline-Kommunikation in Niederspannungsnetzen.
ISBN 978-3-7315-1010-9
- Band 23 Struckmeier, Frederick
Prozesssicherheit von Laserschneidmaschinen – Auflagemessung und Schachtelung
ISBN 978-3-7315-1127-4
- Band 24 Tatzel, Leonie Felica
Verbesserungen beim Laserschneiden mit Methoden des maschinellen Lernens
ISBN 978-3-7315-1128-1

ISSN 2190-6629
ISBN 978-3-7315-1128-1

Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-7315-1128-1



9 783731 511281 >