

# **Verlässliche Identifikation logistischer Entitäten anhand inhärenter visueller Merkmale**

Zur Erlangung des akademischen Grades eines  
**Dr.-Ing.**  
von der Fakultät Maschinenbau  
der Technischen Universität Dortmund  
genehmigte Dissertation

**Jérôme Rutinowski, M.Sc.**  
aus  
Bochum, Deutschland

Tag der mündlichen Prüfung: 25.02.2025

Erste Gutachterin: Prof. Dr.-Ing. Alice Kirchheim  
Zweiter Gutachter: Prof. Dr. Emmanuel Müller

**Dortmund, 2025**



# Vorwort und Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Förder- und Lagerwesen der Technischen Universität Dortmund. Den Rahmen meiner Dissertation bildete die gemeinsame Forschungsarbeit mit dem Fraunhofer-Institut für Materialfluss und Logistik und dem Lehrstuhl für Data Science und Data Engineering der Technischen Universität Dortmund.

Neben meinen Gutachtern Prof. Dr.-Ing. Alice Kirchheim und Prof. Dr. Emmanuel Müller, möchte ich auch Dr.-Ing. Moritz Roidl und Dr.-Ing. Christopher Reining für ihren Rat, ihre Unterstützung und ihr Vertrauen in meine Forschungstätigkeiten danken. Seitens des Fraunhofer-Instituts möchte ich insbesondere meinen Kollegen aus dem Projekt „Silicon Economy Logistics Ecosystems“ dafür danken, dass sie mich inmitten der Covid-Pandemie mit offenen Armen in ihrem Team aufgenommen haben. Ich danke meinen Kollegen am Lehrstuhl für die gute Zusammenarbeit, dabei nicht zuletzt den studentischen und wissenschaftlichen Hilfskräften, die meine Forschung unablässig unterstützt haben.

Auch wenn noch viele weitere Personen zu nennen wären, möchte ich insbesondere Simon Klüttermann und Frederik Polachowski meinen besonderen Dank ausrichten. Beide sind über alle Maße brillante, talentierte und produktive Forscher und nicht zuletzt großartige Kollegen, deren Unterstützung für diese Arbeit unerlässlich war.

Au final, je tiens à remercier ma famille – je n’aurais jamais imaginé atteindre un tel accomplissement sans vous. Chaque réussite est le résultat de votre soutien et je vous dois tout ce que j’ai et tout ce que je suis. Je ferai de mon mieux pour vous rendre ce que vous m’avez offert tout au long de ma vie.

Dortmund, 2025

*Jérôme Rutinowski*



# Kurzfassung

Im industriellen Umfeld und insbesondere in der Logistikbranche ist die Kenntnis über den Aufenthaltsort von Gütern und die Möglichkeit der Identifikation ebener von großem Nutzen. Sofern diese Informationen erhoben werden, geschieht dies durch die Nutzung extrinsischer Merkmale, wie etwa Barcodes. Jedoch sind häufig kostengünstige Entitäten, wie etwa Paletten, bislang nicht serialisiert und können somit nicht über ihren Lebenszyklus verfolgt werden. Die Serialisierung dieser Entitäten anhand ihrer inhärenten visuellen Merkmale bringt jedoch erhebliche Vorteile hinsichtlich der Nachverfolgbarkeit und Prozessoptimierung mit sich. Die dadurch gewonnene Prozesstransparenz wiederum stärkt das Vertrauen in sonst intransparente Vorgänge. Diese Promotionsschrift befasst sich deshalb mit der Bearbeitung der beiden folgenden Forschungsziele:

Das erste Forschungsziel dieser Arbeit ist die Identifikation logistischer Entitäten anhand ihrer inhärenten visuellen Merkmale. Hierbei sollen geeignete logistische Entitäten ausgewählt und entsprechende Datensätze erstellt werden. Anschließend werden verschiedene Identifikationsmethoden auf die Datensätze angewendet und verglichen.

Das zweite Forschungsziel ist die erstmalige Definition und anschließende Quantifikation des Begriffs der Verlässlichkeit im Maschinellen Lernen im Allgemeinen und des Identifikationsverfahrens im Spezifischen. Es wird eine Definition des Begriffs basierend auf seinen Komponenten, die aus der relevanten Literatur deduziert werden, entwickelt und eine geeignete Quantifikationsmetrik erarbeitet.

Zuletzt werden die erstellten Datensätze qualitativ evaluiert und die Identifikationsmethoden anhand ihrer Prädiktionsgenauigkeit bemessen. Es wird weiterhin ein industrienahes Szenario der Identifikation zur Evaluation der Umsetzbarkeit des Verfahrens umgesetzt. Wiederum werden die Definition und Quantifikationsmetrik der Verlässlichkeit anhand ihrer Reliabilität in der Nutzung von Experten evaluiert.



# Abstract

In the industry in general and in logistics in particular, knowledge about the location of goods and the ability to identify them is of tremendous value. If this information is obtained, it is commonly done using extrinsic features, such as barcodes. However, many cost-effective entities, such as pallets, are not serialized and thus cannot be tracked throughout their lifecycle. Serializing these entities based on their inherent visual characteristics would imply significant advantages regarding traceability and process optimization. The resulting process transparency would also increase trust in otherwise opaque operations. This dissertation addresses the following two research objectives:

The first research objective is the identification of logistical entities based on their inherent visual characteristics. This involves selecting suitable logistical entities and creating corresponding datasets. Various identification methods are then applied to these datasets and subsequently compared.

The second research objective is the initial definition and subsequent quantification of the concept of trust in machine learning in general and in the identification process in particular. A definition of the notion, based on its components deduced from relevant literature, is developed, and an appropriate quantification metric is established.

Finally, the created datasets are qualitatively evaluated and the identification methods are assessed based on their prediction accuracy. An industrial identification scenario is implemented to evaluate the feasibility of the approach. Additionally, the definition and quantification metric of trust are evaluated based on their reliability in expert use.



# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b> .....	XII
<b>Tabellenverzeichnis</b> .....	XV
<b>1 Einleitung</b> .....	1
<b>2 Grundlagen und Stand der Technik</b> .....	7
2.1 Aspekte des Materialflusses in der Intralogistik .....	8
2.1.1 Standardisierte Ladungsträger .....	8
2.1.2 Sensorik und Peripheriesysteme .....	10
2.1.3 Digitalisierung und Autonomisierung .....	13
2.1.4 Die Vision des Lagers der Zukunft .....	15
2.2 Identifikationsansätze .....	18
2.2.1 Identifikation anhand extrinsischer Merkmale .....	21
2.2.2 Identifikation anhand inhärenter Merkmale .....	25
2.3 Datensatzerstellung im Maschinellen Lernen .....	32
2.4 Verlässlichkeit als Aspekt des Maschinellen Lernens .....	33
2.5 Wissenschaftliche Beobachtung und Quantifikation .....	36
<b>3 Grenzen bisheriger Forschung</b> .....	41
3.1 Anforderungen an die Identifikation logistischer Entitäten ...	42
3.2 Bestehende Datengrundlage für die Identifikation logistischer Entitäten .....	42
3.3 Bestehende Identifikationsverfahren .....	45
3.4 Verlässlichkeitsbegriff im Maschinellen Lernen .....	48
3.5 Implikationen der Forschungslücke für das entwickelte Vorgehen .....	50
3.6 Abgrenzung des Forschungsgegenstands .....	52

<b>4</b>	<b>Vorgehen zur verlässlichen Identifikation logistischer Entitäten</b>	<b>55</b>
4.1	Erstellung geeigneter Datensätze für die Identifikation logistischer Entitäten	56
4.2	Identifikationsprozess	57
4.2.1	Erstellung, Training und Auswahl der Modelle	57
4.2.2	Beschreibung des Identifikationsprozesses	62
4.2.3	Auswahl der Evaluationsmetriken	63
4.3	Verlässlichkeitsbetrachtung	64
4.3.1	Herleitung der relevanten Aspekte der Verlässlichkeit	65
4.3.2	Evaluation der Verlässlichkeit	75
<b>5</b>	<b>Beschreibung der Datensätze und Experimente</b>	<b>79</b>
5.1	Erstellte Datensätze	80
5.1.1	Datensatz A – pallet-block-502	80
5.1.2	Datensatz B – pallet-block-32965	81
5.1.3	Datensatz C – galvanized-636	83
5.2	Identifikationsexperimente	84
5.2.1	Szenario I – Offline Evaluation (Datensatzversuch)	85
5.2.2	Szenario II – Online Evaluation (Feldversuch)	85
5.3	Verlässlichkeit der Identifikation	93
5.3.1	Definition des Begriffs der Verlässlichkeit	93
5.3.2	Quantifikation der Verlässlichkeit	94
<b>6</b>	<b>Evaluation der Experimente</b>	<b>99</b>
6.1	Datensatzqualität	100
6.1.1	Evaluation von Datensatz A	100
6.1.2	Evaluation von Datensatz B	101
6.1.3	Evaluation von Datensatz C	102
6.2	Identifikationsgenauigkeit	102
6.2.1	Evaluation von Szenario I	102
6.2.2	Evaluation von Szenario II	106
6.3	Verlässlichkeitsuntersuchung	110
6.4	Erkenntnisgewinn und kritische Würdigung der Arbeit	113
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>117</b>
7.1	Erkenntnisse und wissenschaftlicher Beitrag	117
7.2	Forschungsausblick	119
	<b>Literaturverzeichnis</b>	<b>121</b>
	<b>Anhang</b>	<b>149</b>
A.1	Repositories, Datensätze und Dokumentation	149
A.2	Bilder und Icons	149
A.3	Vorveröffentlichungen und Forschungsprojekte	150

A.4 An der Forschungsarbeit beteiligte Personen .....	151
A.5 Studentische Arbeiten im Kontext der Forschungsarbeit .....	153



# Abbildungsverzeichnis

1.1	Grafische Darstellung der Struktur dieser Arbeit. ....	4
2.1	Euro-Palette und Gitterbox als exemplarische Ladungsträger. ....	8
2.2	Exemplarische Taxonomie der Sensorik. ....	10
2.3	Beispiel eines optischen, markerbasierten Motion Capture Systems. ....	12
2.4	Regelkreis eines geläufigen PID-Reglers. ....	14
2.5	Exemplarische Taxonomie der Identifikationsverfahren. ....	19
2.6	Gegenüberstellung der Klassifikation, Detektion und Identifikation. ....	21
2.7	Beispiele extrinsischer Kodierungen: DataMatrix, UPC und QR. ...	21
2.8	Beispiele der Nutzung von Landmarks für die Identifikation. ....	26
2.9	Exemplarische Darstellung eines Embedding Space. ....	27
2.10	Vorgehen bei der Identifikation von Personen. ....	31
2.11	Komponenten der Intersection over Union. ....	31
2.12	Visualisierung der Features eines Klassifikationsmodells. ....	34
3.1	Visualisierung des Part-Based Convolutional Baseline Ansatzes. ...	46
3.2	Visuelle Merkmale einer Palette und einer Stapeltransportkiste. ....	52
4.1	Allgemeiner Workflow der Identifikation in dieser Arbeit. ....	63
4.2	Workflow der Metriken für Szenario II. ....	64
5.1	Auszug des Datensatzes <i>pallet-block-502</i> (Datensatz A). ....	80
5.2	Vereinfachte Visualisierung des Aufnahmesetups für Datensatz A. ....	81
5.3	Vereinfachte Visualisierung des Aufnahmesetups für Datensatz B. ....	81
5.4	Auszug des Datensatzes <i>pallet-block-32965</i> (Datensatz B). ....	82
5.5	Vereinfachte Visualisierung des Aufnahmesetups für Datensatz C. ....	83
5.6	Auszug des Datensatzes <i>galvanized-636</i> (Datensatz C). ....	84
5.7	Workflow der Identifikation für Szenario I. ....	85
5.8	Übersicht der für Szenario II verwendeten Paletten. ....	86
5.9	Anwendungsfall für Szenario II. ....	87

5.10	Workflow der Identifikation für Szenario II. ....	88
5.11	PID-Regelkreis der Drohne für Szenario II. ....	89
5.12	Übersicht der Softwarekomponenten für Szenario II. ....	92
5.13	Operationalisierung des Begriffs und Konzepts der Verlässlichkeit. ....	93
6.1	Visualisierung der Features eines Klotzes aus Datensatz A. ....	103
6.2	Visualisierung der Features eines Klotzes aus Datensatz B. ....	105
6.3	YOLO Bounding Boxes zur Detektion der Klötze und Paletten. ...	106
6.4	Die in Szenario II verwendete Drohne. ....	107
6.5	Flugzeiten und Berechnungszeiten aller 280 Flüge für Szenario II..	108
6.6	Minimale und mittlere Distanzen der 280 Flüge für Szenario II. ...	109
6.7	Sankey Diagramm der Flüge für Szenario II. ....	109
6.8	Beispiel der in Szenario II erfassten Bilder. ....	110
6.9	Die Werte der Aspekte pro Modell und Datensatz. ....	111
6.10	Verlässlichkeitswerte pro Modell und Datensatz. ....	112

# Tabellenverzeichnis

3.1	Identifikations- und Industriedatensätze und ihre Modalitäten. . . .	44
3.2	Benchmarkergebnisse auf dem Datensatz Market-1501. . . . .	47
4.1	PCB Architektur unter Nutzung von ResNet-50 als Backbone. . . .	58
4.2	Features des Knotens des Repräsentationsgraphen. . . . .	59
4.3	Architektur des Siamesischen Graph Neural Network. . . . .	61
4.4	Relevante Literatur zu Fairness im Maschinellen Lernen. . . . .	66
4.5	Relevante Literatur zu Robustheit im Maschinellen Lernen. . . . .	68
4.6	Relevante Literatur zu Integrität im Maschinellen Lernen. . . . .	69
4.7	Relevante Literatur zu Erklärbarkeit im Maschinellen Lernen. . . .	71
4.8	Relevante Literatur zu Sicherheit im Maschinellen Lernen. . . . .	74
5.1	Datensatzaufteilung für das Detektionsmodell YOLO. . . . .	89
5.2	OSD-Skala eines Verlässlichkeitsrisikos des Maschinellen Lernens. . . .	94
5.3	Tabellarische Darstellung der Verlässlichkeitswertbestimmung. . . .	95
5.4	Verlässlichkeitsrisiken des Maschinellen Lernens. . . . .	98
6.1	Ranked Accuracy der Modelle auf Datensatz A. . . . .	104
6.2	Ranked Accuracy der Modelle auf Datensatz B. . . . .	104
6.3	Ranked Accuracy der Modelle auf Datensatz C. . . . .	106
A.1	An der Forschungsarbeit neben dem Autor beteiligte Personen. . . .	152
A.2	Betreute Abschlussarbeiten im Kontext der Dissertation. . . . .	153







# Kapitel 1

## Einleitung

„[...] il n'est pas vray, que deux substances se ressemblent entierement et soyent differentes [...]“

– G. W. Leibniz, „Metaphysische Abhandlung“, Abschnitt IX (1686)

In seinem Diskurs über die Metaphysik proklamierte der deutsche Mathematiker und Philosoph Gottfried Wilhelm Leibniz, frei übersetzt: „Es ist un wahr, dass sich zwei Substanzen gänzlich gleichen und doch unterschiedlich sind.“ [Lei86] Dieses ‚principium identitatis indiscernibilium‘ beschreibt die Unmöglichkeit der Existenz zweier getrennten Entitäten, die in all ihren Eigenschaften identisch sind. Folglich müsste eine jede beliebige Entität von jeder anderen unterscheidbar sein, sofern die Eigenschaften, in denen sie sich unterscheiden, bekannt und messbar sind.

Die Computer Vision und weitere moderne Anwendungen machen sich diese Unterscheidbarkeit zunutze. Unter Betrachtung visueller Unterschiede bieten sich Computer-Vision-Ansätze an, um Entitäten voneinander zu unterscheiden und somit eindeutig zu identifizieren [Ye+21]. So wird die Gesichtserkennung für vielerlei Anwendungen genutzt, beispielsweise in Fällen wie der Entsperrung von Endgeräten [Adj+20]. Den prominentesten Forschungs- und zugleich Anwendungsbereich im Kontext der visuellen Unterscheidung von Individuen stellt die Überwachung des öffentlichen Raums dar. Hierbei werden Passanten anhand ihrer visuellen Charakteristika und anhand weiterer Idiosynkrasien wie ihrem Gangbild identifiziert, verfolgt und mit einer Bilddatenbank abgeglichen [Min+22; NBN19; Wu+19].

Neben Lebewesen weisen auch leblose Entitäten inhärente Charakteristika auf, die sie identifizierbar machen. In Abhängigkeit des Materials oder Herstellungsverfahrens der jeweiligen Entität sind unterschiedlich ausgeprägte Unterscheidungsmerkmale gegeben. Diese reichen von mikroskopischen Unterschieden, bis hin zu welchen, die mit dem menschlichen Auge erfassbar sind [Kla95; TKI17]. Im industriellen Umfeld bietet die Identifikation von Entitäten einen Mehrwert, da sie die Verfolgbarkeit und Prozessanalyse

erheblich beeinflusst. In der Logistikbranche etwa befindet sich eine Vielzahl an Objekten im Umlauf, deren Aufenthaltsort und Bewegung bereits in Teilen nachverfolgt werden [Sha+13]. Dafür werden unter anderem Verfahren verwendet, die extrinsische, artifizielle Merkmale nutzen. Beispiele dafür sind der Barcode, der QR-Code sowie RFID-Systeme [Sha+13].

Motiviert durch die hohen Umlaufmengen und den Bedarf an Interoperabilität werden in der Logistikbranche eine Vielzahl preisgünstiger, standardisierter Entitäten verwendet. Doch sind viele dieser Entitäten nicht serialisiert und dementsprechend nicht entlang ihres Lebenszyklus verfolgbar. Somit werden auf diese Weise keine Informationen aus ihrer Nutzung, ihrem Aufenthaltsort oder ähnlichen Datenpunkten erhoben. Häufig sprechen die geringen Kosten und hohen Mengen der jeweiligen, teilweise als Einweglösungen konzipierten Entitäten gegen eine Serialisierung oder gar die Nutzung extrinsischer Marker. Außerdem ist es für Anwender in der Logistikbranche nicht immer offensichtlich, welchen Mehrwert die Serialisierung einer Entität mit sich brächte. Hinzu kommen Datenschutzbedenken, die einer solch mangelhaften Motivation gegenüberstehen. Insbesondere in kleinen und mittelständischen Unternehmen mit geringen Forschungsetats ist der Mehrwert ohne die nötige technische Kompetenz und strategische Weitsicht, zu gering oder erst weit in der Zukunft greifbar. Einen weiteren Komplikationsfaktor stellt auch die dezentrale Produktionsvergabe dar, wie sie die EPAL e. V. (European Pallet Association) für die Herstellung ihrer Euro-Paletten verwendet. Bei einer solchen Produktionsvergabe wären also im Vorfeld standortübergreifende Serialisierungsmethoden, Datenbanken und Infrastrukturen erforderlich.

Trotz dieser Herausforderungen ist eine Form der Serialisierung von Interesse, da somit eine Repräsentation relevanter Entitäten als Datenpunkte in einem Informationskontinuum ermöglicht wird. Dies erlaubt die Nachverfolgbarkeit der Entitäten und die Messbarkeit von Prozessen, die im Anschluss verbessert werden. Mit dem Wissen über den Aufenthaltsort einer spezifischen Entität, ihres Verwendungszwecks und möglicherweise auch ihrer Relation zu anderen Entitäten im Informationskontinuum ist es möglich, die Lagersteuerung, Disposition und Spedition effizienter zu gestalten. Darüber hinaus generiert das mit der Serialisierung verbundene Wissen eine neue Form der (gegebenenfalls unternehmensinternen) Prozesstransparenz, die sich wiederum vertrauensbildend auswirkt.

Insbesondere das Vertrauen von Nutzern in der Industrie ist von höchster Relevanz, um Akzeptanz für neue Prozesse und Vorgehen zu gewinnen und diese in der Praxis zu etablieren [HE22]. Eine Herausforderung stellt dabei die von Anwendern wahrgenommene Intransparenz moderner Methoden des Maschinellen Lernens dar, die meist als Blackboxmodelle fungieren [Her+23]. Auch wenn die Prozesstransparenz durch die Identifikation logistischer Entitäten gesteigert wird, bleibt doch die unterliegende Metho-

de der Identifikation für die Nutzer intransparent. Aus ebenjenem Grund rückt das Forschungsgebiet des verlässlichen Maschinellen Lernens aktuell in den Fokus, welches sich mitunter mit der Erklärbarkeit von Entscheidungsprozessen befasst [Arr+20]. In diesem noch jungen Forschungsfeld mangelt es jedoch an belastbaren Definitionen und Quantifikationsmetriken, die eine effiziente Kommunikation zwischen sowohl Forschern als auch Nutzern ermöglicht.

Vor diesem Hintergrund beschäftigt sich diese Arbeit mit Methoden der verlässlichen Identifikation logistischer Entitäten unter Verwendung inhärenter visueller Merkmale. Bisher ist die Computer Vision basierte Identifikation logistischer Entitäten unter Verwendung ihrer inhärenten visuellen Merkmale nicht erforscht worden. Das Dissertationsvorhaben ist die Übertragung der Methoden der Computer Vision basierten Personen Re-identifikation auf logistische Entitäten. Dies hat die Identifikation, ohne Veränderung der Entitäten durch artifizielle Identifikatoren zum Ziel. Auf diese Weise bleiben Produktionsvorgänge und -kosten unverändert und die Transparenz und damit das Vertrauen in unternehmensweite Abläufe werden gesteigert.

Die Herausforderung bei dem Unterfangen liegt darin, hohe Genauigkeiten im Verfahren zu gewährleisten, während von kostenintensiven und komplexen Verfahren abgesehen wird. Stattdessen werden mit dem bloßen Auge erkennbare Oberflächenstrukturen genutzt und von gängigen Kamerasystemen erfasst. Als mögliche Entitäten werden initial Objekte aus Holz und galvanisiertem Stahl verwendet. Bei den Objekten handelt es sich um weit verbreitete Objekte aus der Logistikbranche.

Weiter ist es das Ziel dieses Dissertationsvorhabens erstmalig eine operationelle Definition und Quantifikationsmetrik für den Begriff der Verlässlichkeit im Maschinellen Lernen zu erarbeiten. Bislang wurde die Verlässlichkeit von Modellen und Datensätzen des Maschinellen Lernens noch nicht vollumfänglich definiert und quantifiziert. Aus diesem Grund wird sowohl eine Definition des Begriffs der Verlässlichkeit sowie eine Quantifikationsmetrik der Verlässlichkeit erarbeitet und auf einige der hierin erarbeiteten Identifikationsmethoden angewandt.

Konkret verfolgt diese Arbeit somit die folgenden Forschungsziele: Das erste Ziel sind für die Identifikation logistischer Entitäten geeignete Datensätze. Dafür werden adäquate logistische Entitäten ausgewählt und strukturiert Datensätze erstellt. Das zweite Ziel sind Methoden zur Identifikation logistischer Entitäten. Zum Vergleich der verwendeten und erarbeiteten Methoden werden im Vorfeld geeignete Evaluationsmetriken ausgewählt. Das dritte Ziel ist eine Definition des Begriffs der Verlässlichkeit im Kontext des Maschinellen Lernens. Diese Definition ist sowohl für die Anwendung auf die genutzten Daten, Kontextinformationen als auch Modelle zu entwickeln. Zuletzt ist das vierte Ziel der Arbeit eine Quantifikationsmetrik der Verlässlichkeit. Mit dem Fokus auf die Identifikation im industriellen Kontext gilt

es somit eine Möglichkeit zu erforschen, Aspekte der Verlässlichkeit der erarbeiteten Identifikationsmethoden zu quantifizieren.

Um die oben genannten Forschungsziele zu erreichen, gilt es zwei zentrale Forschungsfragen (FF) zu beantworten:

FF1 *Wie können logistische Entitäten anhand ihrer inhärenten visuellen Merkmale identifiziert werden?*

FF2 *Wie kann der Begriff der Verlässlichkeit im Kontext der Identifikation definiert und quantifiziert werden?*

Zusammenfassend, gilt es in dieser Arbeit ein Vorgehen zur Identifikation logistischer Entitäten anhand ihrer inhärenten visuellen Merkmale zu beschreiben. Dabei werden zunächst geeignete logistische Entitäten ausgewählt und passende Datensätze erstellt, auf die im Anschluss Identifikationsverfahren angewandt und miteinander verglichen werden. Weiter werden geeignete Arten der Datenrepräsentation ausgewählt und eine Definition und Quantifikationsmetrik hinsichtlich der Verlässlichkeit der Identifikation im Maschinellen Lernen erarbeitet.

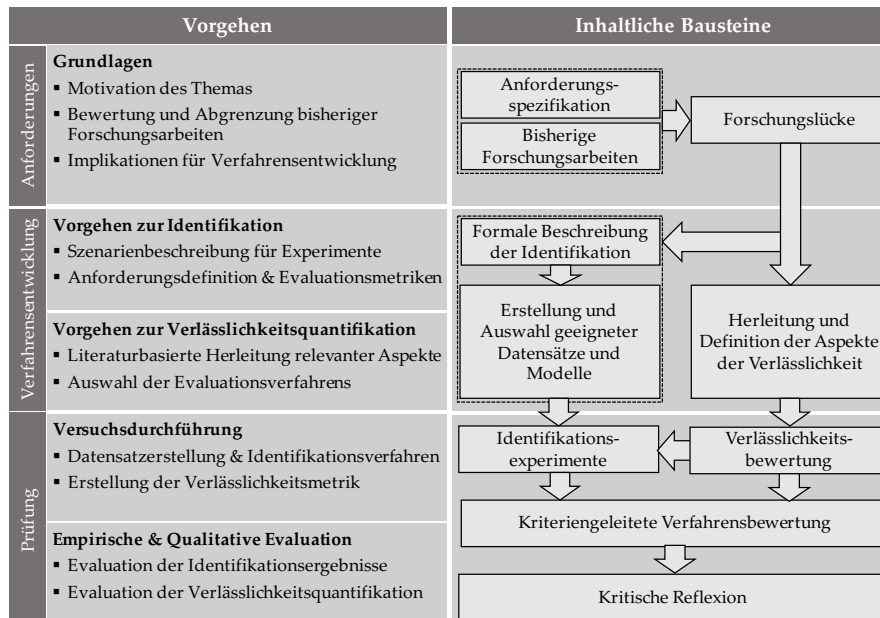


Abb. 1.1 Grafische Darstellung der Struktur dieser Arbeit.

Die weitere Arbeit gliedert sich wie folgt:

In Kapitel 2 wird zunächst ein einführender Überblick über die relevanten Forschungsgrundlagen sowie den Stand der Technik gegeben. Es wird außerdem ein Einblick in den Stand der Technik hinsichtlich des Materialflusses in der Intralogistik geboten. Basierend auf einer umfassenden Litera-

turrecherche werden insbesondere Methoden der Identifikation sowie der Forschungsstand bezüglich der Verlässlichkeit Maschinellem Lernverfahren untersucht.

In Kapitel 3 werden die Grenzen der relevanten Forschung diskutiert. Aufbauend auf diesen Grenzen werden die Vision der Identifikation logistischer Entitäten beschrieben und damit einhergehende Anforderungen aufgezeigt.

In Kapitel 4 wird das eigens entwickelte Vorgehen zur Identifikation logistischer Entitäten und der initiale Rahmen der Verlässlichkeitsbetrachtung vorgestellt. Dabei werden die relevanten Aspekte der Verlässlichkeit aus der Literatur hergeleitet und der vorgeschlagene Identifikationsprozess formal beschrieben.

In Kapitel 5 werden die erstellten Datensätze und experimentellen Szenarien für die Identifikation sowie das Vorgehen zur Evaluation der Verlässlichkeit dargestellt.

In Kapitel 6 werden abschließend die Ergebnisse dieser Arbeit präsentiert und evaluiert. Konkret werden dabei die Datensatzqualität, die Identifikationsgenauigkeit und die erarbeitete Verlässlichkeitsmetrik untersucht.

In Kapitel 7 werden die Erkenntnisse dieser Arbeit zusammenfassend reflektiert und ein Forschungsausblick angeboten. Die Struktur der Arbeit ist Abb. 1.1 zu entnehmen.



## Kapitel 2

# Grundlagen und Stand der Technik

**Zusammenfassung** Das zweite Kapitel dieser Arbeit legt die weiterführend genutzte Terminologie sowie die Grundlagen des relevanten Forschungsstands dar. Zuerst werden die relevanten technischen Aspekte der Intra-logistik, besonders des Materialflusses der genutzten Ladungsträger erläutert. Es werden weiterhin gängige Methoden und Anwendungsfelder der Identifikation, inklusive gängigerweise genutzter Evaluationsmetriken präsentiert. Dabei wird methodisch in die Nutzung extrinsischer und inhärenter Merkmale unterschieden. In diesem Kontext wird auch auf die Erstellung geeigneter Datensätze eingegangen. Weiter werden der Begriff und die Charakteristika der Verlässlichkeit im Kontext des Maschinellen Lernens dargestellt.

„If I have seen further it is by standing on the shoulders of giants.“  
– Isaac Newton, Brief an Robert Hooke (1675)

## 2.1 Aspekte des Materialflusses in der Intralogistik

Der erste Abschnitt dieses Kapitels beschäftigt sich mit den für diese Arbeit relevanten Aspekten der Intralogistik, insbesondere hinsichtlich des Materialflusses. Dabei wird die Relevanz standardisierter Ladungsträger und die Nutzung von Sensorik und Peripheriesystemen zur Verfolgung ebenjener beleuchtet. Darüber hinaus werden bestehende Schritte hin zur Digitalisierung der Intralogistikprozesse vorgestellt. Zuletzt wird die historische Vision der Forschungsgemeinde hinsichtlich der Gestaltung eines Lagers der Zukunft untersucht.

### 2.1.1 Standardisierte Ladungsträger

In der Industrie im weiteren und in der Logistik im engeren Sinne werden diverse standardisierte Objekte verwendet. Dabei wird der Begriff der Standardisierung als konsensbasierte Vereinheitlichung verstanden, welche die Lösung eines wiederkehrenden Problems ermöglicht [Tas00]. Die Subjekte einer Standardisierung sind dabei unterschiedlicher Natur – beispielsweise Bauteile, Maßeinheiten oder Prozesse.

Einer der Vorteile der Standardisierung liegt in der durch sie gewonnenen Interoperabilität zwischen unterschiedlichen Unternehmen und ihren Peripheriegeräten, wie etwa Laderampen oder Förderbändern [Muc+18]. Weiter ermöglicht eine Standardisierung die Modularität von Komponenten und die Verringerung von Produktionskosten [Muc+18]. In der Logistik betrifft die Standardisierung beispielsweise Ladungsträger.



**Abb. 2.1** Handelsübliche Euro-Palette (a) und Gitterbox (b) als exemplarische Ladungsträger.

Ladungsträger fassen gleiche oder ungleiche Güter für den Transport zu einer Ladeinheit zusammen und schützen die Güter teilweise während des Transports [Gud05, S. 413]. Es wird eine Unterscheidung von Ladungsträgern hinsichtlich ihrer Unterfahrbarkeit getroffen [Ros16]. Unterfahrbare Ladungsträger besitzen eine tragende und umschließende Plattform, wie etwa Gitterboxen [DIN 15155:1986-12] und Paletten [DIN 13698-1:2003-1] (siehe Abb. 2.1). Beispiele für nicht unterfahrbare Ladungsträger sind Klein-

ladungsträger [DIN 13199-1:2000-1] aus Polymer oder Stapeltransportkisten aus galvanisiertem Stahl.

So unverzichtbar Ladungsträger auch sind, stellen sie doch gleichzeitig eine Form der Kapitalbindung dar. Aus diesem Grund ist die intensive und insbesondere beladene Nutzung standardisierter, identifizierbarer Ladungsträger erstrebenswert [Gud05, S. 88]. In der Regel werden entweder die Ladungsträger oder die gesamte Ladeinheit (Ladungsträger, Sicherungsmittel und Transportgut) über extrinsische Merkmale (siehe Kapitel 2.2.1) identifizierbar gemacht [Sha+13]. In vielen Fällen findet jedoch eine Nutzung von Ladungsträgern ohne jegliche Identifikation statt. Dies liegt oftmals am Mehraufwand und den zusätzlichen Kosten, die mit der Nutzung jeglicher extrinsischer Identifikationsmethode verbunden sind [SH11]. Insbesondere bei Transporthilfsmitteln, die geringe Produktionskosten aufweisen, ist diese Entscheidung naheliegend. Über die bereits genannte Interoperabilität hinaus ermöglichen einige Ladungsträger die Implementierung von Tausch- und Pfandsystemen. Dank Standardisierung ist dies unter anderem bei der Nutzung von CHEP- und EPAL-Paletten oder Gitterboxen der Fall [KLZ05]. Während der Nutzung ist es wichtig, die Unversehrtheit der Ladungsträger zu überprüfen und diese gegebenenfalls auszutauschen oder zu reparieren. Auch hierfür gibt es dank Standardisierung definierte Kriterien, die zur sicheren Nutzung der Ladungsträger beitragen. Ein Beispiel dafür sind die Kriterien, die von der EPAL hinsichtlich ihrer Paletten und Gitterboxen bereitgestellt werden [KLZ05].

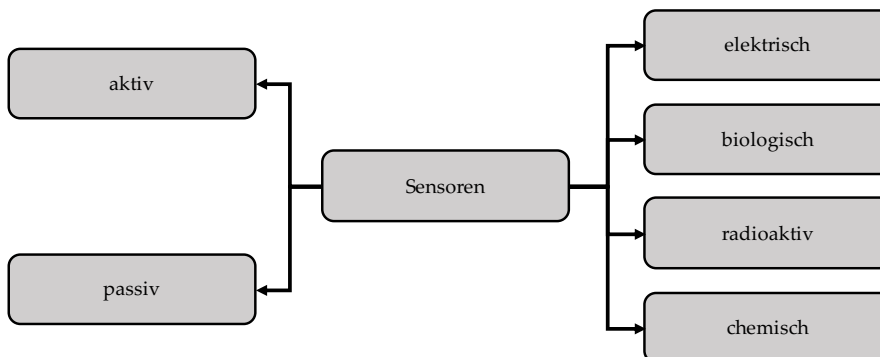
Standardisierte Ladungsträger sind wesentliche Instrumente zur Effizienzsteigerung und Kostenreduzierung in der Logistik. Sie werden aus diesem Grund in diversen Bereichen der Logistik eingesetzt. In der Intralogistik sind Ladungsträger unabdinglich. Standardisierte Ladungsträger bestimmen die Auslegung von Fördersystemen, die entsprechend dimensioniert sind. Beispiele hierfür sind Stetig- und Unstetigförderer (zum Beispiel Handgabelhubwagen und Gabelstapler) sowie Lagersysteme (zum Beispiel Regalsysteme und Förderbänder). Die standardisierte Nutzung von Ladungsträgern ermöglicht in diesem Kontext eine effiziente Lagerverwaltung und Kommissionierung sowie eine Optimierung des innerbetrieblichen und Warenflusses und Versands [PF23; Vil+19].

In der Produktion sind Ladungsträger essentiell, die auf Montagevorrichtungen und Förderbänder abgestimmt sind und somit eine effiziente Material- und Werkstückversorgung ermöglichen. Hinzu kommt die Relevanz der Verlässlichkeit hinsichtlich der Qualität und Materialeigenschaften der Ladungsträger, die dank Standardisierung gewährleistet wird. So sind etwa Eigenschaften wie die Hitzebeständigkeit oder die Traglast eines Ladungsträgers in einigen Produktionsumgebungen ausschlaggebend [Hen+14; SP07].

Beim Verlassen eines Unternehmens spielt die Standardisierung von Ladungsträgern ebenfalls eine große Rolle. Im Bereich der Transportlogistik sind beispielsweise ISO-Container, Paletten und Wechselbrücken für den Transport über Luft, Wasser und Straße unerlässlich. Diese standardisierten Ladungsträger erleichtern den Transport nicht nur durch das Zusammenfassen von Gütern. Ihre standardisierten Maße sind ebenfalls entscheidend für die Beladungsplanung von LKW, Schiffen und Flugzeugen, wodurch sie zur Planungssicherheit und Effizienz im globalen Warenverkehr beitragen [Yur+21].

### 2.1.2 Sensorik und Peripheriesysteme

Sensoren sind Geräte oder Module, die Ereignisse oder Zustandsänderungen erfassen und die daraus resultierenden Informationen in maschinenlesbare Formate umwandeln [Jav+21]. Aufgrund dieser Eigenschaft sind Sensoren für jegliche Ausprägungen der Digitalisierung und Autonomisierung unerlässlich. Die Möglichkeit der lückenlosen Erfassung der Vorgänge in einem System werden als eine der Grundlagen der Industrie 4.0 erachtet und stehen somit für eine weitreichende innerbetriebliche Digitalisierung [Sch+20, S. 18]. Eine Taxonomie der Sensorik ist basierend auf unterschiedlichen Diskriminierungscharakteristika arrangierbar. So werden Sensoren beispielsweise dahingehend unterschieden, ob sie aktiv oder passiv Signale aufnehmen [Jav+21]. Beispiele für aktive Sensoren stellen Radar oder Lidar dar, die Signale emittieren, die sie anschließend wieder aufnehmen. Im Gegensatz dazu stehen passive Sensoren, wie RGB-Kameras, die lediglich unter Aufnahme der Signale funktionieren, die ihnen extern zur Verfügung stehen.



**Abb. 2.2** Exemplarische Taxonomie der Sensorik anhand der Signalaufnahme (links) und der Ereignisdetektion (rechts).

Eine alternative Unterscheidung der Sensortechnologien stellt die Unterscheidung hinsichtlich der Art der Ereignis- oder Zustandsänderungsdetektion dar [Jav+21]. In diesem Kontext wird etwa zwischen der Nutzung

elektrischer, chemischer, biologischer oder radioaktiver Detektion unterschieden. Die Methode, mit welcher dieser Detektionsimpuls in ein elektrisches Signal umgewandelt wird, sei es photoelektrisch, thermoelektrisch, elektrochemisch, elektromagnetisch oder thermooptisch, stellt eine weitere Unterscheidungsweise dar [Jav+21]. Eine exemplarische Übersicht der Taxonomie der Sensortechnologien ist Abb. 2.2 zu entnehmen.

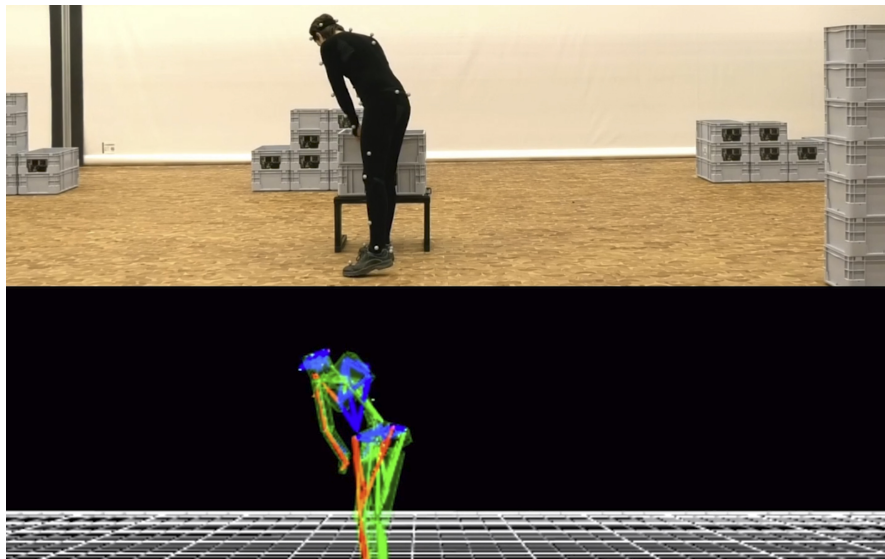
RGB-Kameras sind eine weitläufig genutzte Art der Sensorik. Die Auswertung zweidimensionaler Bilder mit drei Farbkanälen (Rot, Grün, Blau) ist seit langem und auf verschiedene Wege möglich [Kla95; Vou+18]. In der letzten Dekade ist die Nutzung von Deep Learning Methoden dafür besonders populär geworden, insbesondere unter Verwendung von Convolutional Neural Networks (CNN) [For+99; Li+21]. CNN haben sich dank der Nutzung des namensgebenden Convolutional Layers in Kombination mit Pooling Layern, als eine effektive und zeiteffiziente Methode der Bildverarbeitung herausgestellt.

Neben der Verwendung gängiger RGB-Kameras hat sich auch die Nutzung von Tiefenkameras für einige Anwendungsfälle etabliert. Tiefenkameras oder RGBD-Kameras (wobei das *D* für Depth steht) sind Kameras, die Punktwolken ausgeben und somit dreidimensionale Bildinformationen festhalten. Dabei wird für die Erstellung der Punktwolke zwischen der Nutzung von Stereo-Kameras, dem Structured-Light Verfahren und dem Time-Of-Flight Verfahren unterschieden.

Stereo-Kamerasysteme nutzen zwei Kameras, welche die gleiche Szene aus leicht versetzten Perspektiven betrachten [HH22; KK99, S. 269]. Durch die Position und Rotation einer Kamera relativ zur anderen wird die 6D-Position einzelner Punkte translatorisch und rotatorisch bestimmt. Dabei ist es entscheidend, einen Punkt, dessen Position ermittelt wird, in beiden Kamerabildern als denselben zu identifizieren [KK99; SS02, S. 270]. Die Herausforderung besteht darin, Punkte in homogenen Gruppen zu unterscheiden. Die Homogenität spiegelt sich hierbei wider, durch Bereiche gleicher Farbe, welche repräsentativ für die Tiefeninformation sind. Dies wird durch das natürliche Rauschen von Infrarot-Licht in der Umgebung unterstützt [HH22]. Eine weitere Möglichkeit besteht darin, diese Erkennung durch aktive Infrarot-Projektoren zu verbessern, die in der Kamera integriert sind [HH22].

Bei der Nutzung von Time-Of-Flight Systemen wird anstelle einer zweiten Kamera eine Lichtquelle verwendet, häufig in Form eines Laser-Emitters [FAT11; HH22]. Diese Lichtquelle beleuchtet schrittweise die Szene, während die Kamera die Zeit misst, die das ausgestrahlte Licht benötigt, um ihren Photodetektor zu erreichen [FAT11]. Die Zeitdifferenz zwischen dem Aussenden und dem Eintreffen des Lichts sowie die bekannte Lichtgeschwindigkeit ermöglichen die Bestimmung der Entfernung eines Punktes relativ zur Kamera in der Punktwolke [FAT11].

Ähnlich wie bei Time-Of-Flight Kameras verwendet auch das Structured-Light Verfahren (auch Streifenprojektion genannt) nur eine Kamera. Zusätzlich wird ein Structured-Light Projektor in unmittelbarer Nähe der Kamera positioniert [Gen11]. Beispiele für solche Projektoren sind Medienprojektoren oder Lichtquellen, die Licht in einem festen Muster abgeben. Bei diesem Muster handelt es sich oft um einen einzelnen Strich, der längs über die erfasste Szene projiziert wird [Gen11]. In beiden Fällen wird das Wissen über das ausgesendete Licht genutzt, um anschließend Informationen aus den von der Kamera aufgezeichneten Bildern zu extrahieren [Gen11; HH22]. Um eine Punktwolke zu erstellen, ist in jedem Fall die Aufnahme und Analyse einer Serie solcher Bilder erforderlich [HH22].



**Abb. 2.3** Beispiel eines optischen, markerbasierten Motion Capture Systems in der Forschungshalle des Lehrstuhls für Förder- und Lagerwesens der Technischen Universität Dortmund.

Alternative Verfahren mit ähnlichen Funktionsweisen, beispielsweise unter Nutzung von Laser [Lat15; Shi+16] oder Lidar [FMD19; SCK16], sind ebenfalls zu nennen. Eine Sensorik, die sich von den bisher genannten unterscheidet, stellen Motion Capture Systeme dar (siehe Abb. 2.3). Diese Systeme ermöglichen die Verfolgung von Objekten und Personen, beispielsweise mittels retroreflektiver Marker [Fur+19, S. 16]. Diese reflektieren Licht direkt zur Infrarotlichtquelle zurück, wodurch die genaue Verfolgung der Position der Marker ermöglicht wird. In Kombination mit mehreren Kameras, welche die Marker im Raum aufzeichnen, ermöglicht ein Motion Capture System die Generierung hochpräziser dreidimensionaler Bewegungsdaten. Alternative Ansätze sowohl optischer als auch nicht-optischer Natur existieren eben-

falls [ZH08]. Motion Capture Systeme werden häufig in der Unterhaltungsindustrie für Animationen in Filmen oder Videospielen [Sha+19] sowie in biomechanischen Studien und Sportanalysen eingesetzt [KR18]. Eine weitere Anwendung von Motion Capture Systemen stellt aber auch die menschliche Aktivitätserkennung im Anwendungsfall von logistischen Prozessen dar [Nie+20]. Motion Capture Systeme ermöglicht es somit feinste Bewegungsdetails aufzuzeichnen und dient sowohl kreativen als auch wissenschaftlichen Anwendungen, in denen präzise Bewegungsdaten von entscheidender Bedeutung sind. In Kombination mit Bilddaten werden Motion Capture Systeme für die präzise Verfolgung und teil-automatisierte Annotation im Bereich des Trackings und der Pose Estimation eingesetzt [You+24].

### 2.1.3 Digitalisierung und Autonomisierung

Sensoren und ihre Peripheriesysteme spielen die kritische Rolle eines Signalgebers. Sie liefern sie Daten, welche die Basis der Digitalisierung und Automatisierung logistischer Prozesse darstellen. Beispielsweise werden mittels solcher Signale, Informationen über Anlagen und Systeme übermittelt, wodurch die prädiktive Analyse relevanter Zustandsgrößen von Gerätschaften ermöglicht wird. Mittels dieser Informationen werden beispielsweise Einblicke über Abnutzungsgrade von Maschinenbauteilen inferiert, welche die prädiktive Wartung von Gerätschaften vor ihrem Ausfall ermöglichen [Zon+20].

Darüber hinaus, steuern und optimieren Warehouse Managementsysteme bestehende Lagerprozesse, durch den Erhalt von Informationen, beispielsweise mittels RFID-Transpondern oder Barcodes [RSR12]. In der Transportlogistik wiederum sind Entwicklungen wie der elektronische Frachtbrief relevante Digitalisierungsmaßnahmen [BS22; GS23].

Über die reine Digitalisierung hinaus, tragen Sensoren zur Autonomisierung industrieller Robotik bei. Mit Autonomisierung im Kontext der Robotik ist hierbei die Befähigung eines Agenten zur (partiellen) eigenständigen Ausführung einer Aufgabe gemeint, die zuvor von einem Menschen ausgeführt worden wäre [PSW00]. Dabei wird hinsichtlich der Eigenständigkeit in unterschiedliche Autonomiegrade differenziert, wie der autonomen Informationsakquisition und -verarbeitung, der autonomen Interpretation dieser Informationen und der darauf basierenden Entscheidungsfindung, sowie der autonomen Ausführung einer Handlung basierend auf der Informationsaufnahme und -interpretation [SHW07; Vel02].

In diesem Kontext fungieren die relevanten Informationen als Sensorsignale und dienen dementsprechend als Eingangsgrößen für Regel- und Steuerungssysteme, die wiederum Ausgangssignale erzeugen. Konkreter werden dabei Regler verwendet, die Roboter und Anlagen basierend auf Eingangssignalen steuern. Dieser Ansatz wird bereits seit langem verfolgt.

Ein Beispiel dafür ist die Nutzung klassischer Regler [Slo88], um rudimentäre Robotik und Anlagen zu steuern.

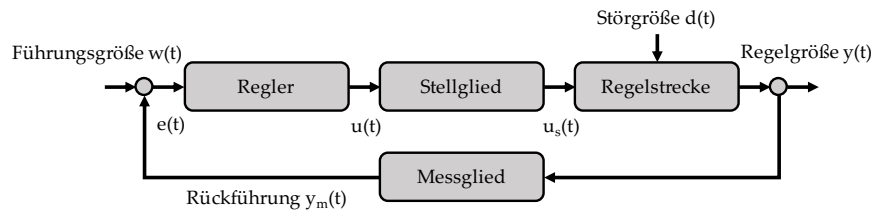


Abb. 2.4 Regelkreis eines geläufigen PID-Reglers.

Dieser Ansatz wurde über die Jahre verfeinert, etwa unter der Nutzung von Fuzzy Logik [Kon+19], um beispielsweise Roboter zu steuern. Ein Beispiel eines solchen Reglers ist der PID-Regler (proportional-integral-derivative) [ACL05], wie er in Abb. 2.4 dargestellt ist. Der PID-Regler setzt sich aus drei Anteilen zusammen:

- Proportionaler Anteil (P): Der proportionale Anteil reagiert proportional zur Regelabweichung. Er sorgt dafür, die Regelgröße direkt verhältnismäßig zur Regelabweichung zu verändern, um das System in Richtung des Sollwerts zu steuern.
- Integraler Anteil (I): Der integrale Anteil verringert den stationären Fehler, indem er niedrigfrequente Kompensation durch einen Integrator ermöglicht. Durch die Berücksichtigung der vorangegangenen Regelabweichung trägt der I-Anteil dazu bei, stetige Fehler im System auszugleichen.
- Differenzieller Anteil (D): Der differenzielle Anteil verbessert die transienten Antwortzeiten, da er hohe Frequenzen kompensiert. Dieser Anteil berücksichtigt die Änderungsrate der Regelabweichung und trägt dazu bei, die Regelung zu verlangsamen oder zu beschleunigen, um unerwünschte Oszillationen oder Überschwingen zu minimieren.

Die PID-Regelungsgleichung kombiniert diese Anteile, um die Stellgröße  $u(t)$  zu berechnen, die auf das zu regelnde System wirkt:

$$u(t) = K_p e(t) + K_i \int (e(t) dt) + K_d \frac{de(t)}{dt} \quad (2.1)$$

Die Gewichtungsfaktoren  $K_p$  (proportional),  $K_i$  (integral) und  $K_d$  (differenziell) der Eingangsgröße  $e(t)$  werden entsprechend eingestellt, um eine optimale Leistung des Regelkreises zu erreichen.

In der Umsetzung ermöglicht die Nutzung von Regelsystemen den Betrieb von Robotik, wie sie in der Industrie vielfach verwendet wird. Weitläufig eingesetzte Beispiele dafür sind Shuttle Systeme für die automatische Bedienung von Hochregallagern [Fer+19; Zha+20], Knickarmroboter in Kombination mit

Computer Vision für die Aufgabe des Bin Pickings [Gou+21; TKW10], mobile Robotik wie fahrerlose Transportfahrzeuge [THH09; Viv+10] oder autonome mobile Roboter wie Loadrunner für den Transport von Gütern innerhalb eines Lagers [Hom+20; Kam+11], oder Drohnen für die Inventarisierung und Inspektion von Gütern [WMN19; XKM18].

#### 2.1.4 Die Vision des Lagers der Zukunft

In ihrer Gesamtheit repräsentieren die in den vorangegangenen Abschnitten vorgestellten Technologien und Komponenten einige der Ergebnisse des fortwährenden Versuchs, Prozesse zu standardisieren und zu optimieren. So ist es seit jeher ein Ziel der Forschung, den Status Quo durch Erkenntnisgewinne zu verändern. Das Ziel der industriellen Anwender und ebenjener, die den Transfer zwischen Forschung und Industrie zu ermöglichen versuchen, ist es wiederum aus den Erkenntnissen der Forschung relevante und profitable Neuentwicklungen zu schaffen und sie in der Industrie zu implementieren. Dieser stetige Wandel ist auch der Intralogistik nicht fremd und so arbeiten Logistikforschung und Industrie kontinuierlich an der Optimierung und funktionellen Erweiterung bestehender Logistikhösungen und -prozesse. In diesem Prozess haben Forscher vielfach eine Vision eines ‚Lagers der Zukunft‘ formuliert. Diese Vision ist folglich über die Jahre und Jahrzehnte im Einklang mit den technologischen Möglichkeiten gewachsen und hat sich progressiv weiterentwickelt. Dabei wurde im Rahmen der Logistikforschung stets versucht, die Entwicklung des Lagers – und damit das Lager der Zukunft – vorherzusagen.

Zunächst einmal ist der Prozess des Lagerns definiert als „[...] das Aufbewahren und Bereithalten der Bestände einer Anzahl von Artikeln“ [Gud04, S. 579]. Das Lager ist aus Sicht der Architekturwissenschaften wiederum definiert als ein Gebäude, das für die Unterbringung von Gütern unterschiedlicher Art entworfen wurde [Har06]. Das womöglich erste dokumentierte und als Lager genutzte Gebäude wird auf das zweite Jahrhundert vor Christus in Ostia, Italien, datiert [Ric71]. Am Munde des Flusses Tiber und nahe der Stadt Rom gelegen, handelte es sich dabei vermutlich um ein Hafenerlager des antiken römischen Reichs.

In der Zeit der Industrialisierung gewann das Lager über den militärischen Einsatz hinaus an Relevanz, wobei das Lager im Rahmen der Forschung anfänglich geringgeschätzt wurde [McK83]. In den 1920er Jahren äußerten Forscher Bedenken und Beobachtungen bezüglich der Einführung von Sicherheitsstandards, um schwerwiegende Unfälle zu vermeiden [Ada20]. Somit beinhaltet diese Vision des Lagers der Zukunft, die Standardisierung von Bau- und Sicherheitsmaßnahmen. Bis in die 1970er Jahre hatten beachtliche Entwicklungen hinsichtlich der Automatisierung und Robotik stattgefunden, unter anderem als Teil der Forschung für die Rüstungsindustrie. Es

lag somit nahe, diese Entwicklungen auch in Lagerumgebungen zu integrieren. Entgegen der heutigen Inklination ebenjenes zu tun, gab es in der Forschungsgemeinde einige Zweifel an der Sinnhaftigkeit der Automatisierung des Lagers durch den Einsatz von Robotik [Bar77; Bla79]. In etwa demselben Zeitraum wuchs in der Industrie die Aufmerksamkeit, die der Lagerplanung und -forschung geschenkt wurde [McK83]. McKinnon et al. wiesen dennoch im Jahr 1983 darauf hin, das Ansehen der Lagerforschung verdiene weiterhin einen höheren Stellenwert [McK83]. Ebenfalls wandte sich der Blick der Forschungsgemeinde in den 1980er Jahren auf die Simulation des Lagers. Dieser wurde großes Potential für den zukünftigen Einsatz zugeschrieben. Es wurde allerdings auch angemerkt, das Lager sei ein zu komplexes System und erfordere zu hohe Rechenleistungen, um vollständig simuliert zu werden [AG85].

In seinem umfassenden Werk „Reinventing the Warehouse“ von 1993, stellt Roy L. Harmon seine Vorstellung eines Lagers (und der Transportlogistik) der Zukunft vor. Der Autor beschreibt dabei modulare Lager, die teilweise automatisiert sind. Er betrachtet jedoch die Automatisierung mit Vorsicht, da sie für manche Aufgaben besser geeignet sei als für andere [Har93, S. 89]. Darüber hinaus sei die Größe des Lagers in Zukunft zu reduzieren, da hier ein ‚just-in-case‘ Ansatz verfolgt worden sei, der kostspielig und ineffizient sei [Har93, S. 90]. Der Autor nennt außerdem die Nutzung von Computersystem für den Austausch und das Festhalten von Daten in digitaler Form als eine Chance für die Gestaltung eines modernen Lagers [Har93, S. 159]. Im Jahr 1996 wurde womöglich zum ersten Mal der Vorschlag unterbreitet, im Lager der Zukunft zellulare, autonome Maschinen zu nutzen [Sak+96]. Darunter wurde demnach autonome, modulare Robotik, wie beispielsweise Fördertechnik verstanden, die dem Lager eine zusätzliche Flexibilität und Adaptivität ermöglicht.

Kurz vor dem Millenniumswechsel formulierte Thompson Brockman, Forscher des Institute of Industrial and Systems Engineers (IISE), die 21 Warehousing Trends [Bro99]. Diese Trends würden laut Brockman das Lager der Zukunft gestalten [Bro99]. Unter anderem nennt er den vermehrten Einsatz von Informationstechnologie und Automatisierungstechnik, die Nutzung des Internets und die Verminderung der Nutzung des Papiers, hin zu einem papierlosen Lager. Dadurch erhoffte sich Brockmann die Verringerung des Zeitaufwands von Prozessen und des Platzbedarfs, die Ermöglichung von Value-Added Services sowie eine erhöhte Messbarkeit der Lagerprozesse [Bro99].

Kurz nachdem der Begriff des Internets der Dinge 1999 von Kevin Ashton erstmals artikuliert [Ash09] wurde, prägte Professor Michael ten Hompel den Begriff im Kontext der Intralogistik [BH07; Hom06]. In seinen Werken beschreibt er die Potenziale der Autonomisierung im Kontext der Intralogistik. Konkret werden dabei die Kombination von RFID-Transpondern

und Agentensoftware zu einem ‚Aware Object‘ genannt, die im Internet der Dinge untereinander kommunizieren und Ressourcen anfordern [Hom06, S. 272]. Mit der Einführung solcher Konzepte wurde erstmals eine konkrete und technisch umsetzbare Möglichkeit in Aussicht gestellt, intralogistische Prozesse digital nachzuverfolgen und zu autonomisieren. Gemeinsam mit Professor Willibald Günthner veröffentlichte Professor ten Hompel im Jahr 2010 ein weiteres Werk zu diesem Thema, das ebenjene Konzepte noch einmal aufgreift und tiefergehend beschreibt [GH10]. In diesem Werk wird der Vision des Internets der Dinge in der Intralogistik noch einmal ein separates Kapitel gewidmet, in welchem das Konzept wie folgt resümiert wird: „[Es entsteht] ein Materialflusssystem, das dem Prinzip nach wie ein Computernetzwerk funktioniert: Intelligente Infrastruktur stimmt sich mit Transporteinheiten ab und realisiert ein dezentrales, kooperatives und beliebig veränderbares Transportnetzwerk.“ [GH10, S. 43]

Im Jahr 2014 formulierten die Professoren Michael ten Hompel und Michael Henke ihre Vision des idealen logistischen Raums (und damit des idealen Lagers) und verlautbarten: „Der ideale logistische Raum ist leer!“ [HH14] Im Zentrum dieser Vision der Logistik 4.0 steht die Beseitigung von nahe zu jeglicher stationärer Fördertechnik. Im Gegensatz dazu sieht diese Vision die Nutzung autonomer Fahrzeuge und Fahrzeugswärme vor. Die genutzten Regale und Ladungsträger im Lager werden durch eine Vernetzung mit dem restlichen System zu Teilen des cyberphysischen Systems und ergänzen somit die innerbetriebliche Informationsweitergabe [HH14]. Ergänzend dazu fügte Moritz Roidl in seiner Dissertationsschrift an, der leere Raum sei „[...] Ausgangszustand und zugleich Endzustand der logistischen Aktivität“, womit er darauf hinweist, wenn ein Lager durchweg keine Objekte enthielte, so stünde „[...] dieser Zustand im Widerspruch zur Leistungserbringung des logistischen Systems, für die möglichst viele Objekte in kurzer Zeit hindurchbewegt werden müssen, also der Raum eigentlich möglichst gefüllt sein sollte.“ [Roi22] Im Gegensatz zu der bereits genannten Skepsis gegenüber der Automatisierung des Lagers, sprechen sich Esmailian et al. im Jahr 2016 für die weitere Automatisierung in zukünftigen Lagerplanungen aus [EBW16]. Die Forscher sehen die Automatisierung als unausweichliche Konsequenz der Einführung der just-in-time und Lean Manufacturing Konzepte, die eine erhöhte Flexibilität und wiederkehrende Neukonfiguration des Lagers erfordern [EBW16]. Aufbauend auf den beiden vorangegangenen Notionen von ten Hompel et al. und Esmailian et al., haben Thomeier et al. für das Unternehmen Swisslog eine Veröffentlichung zur Planung des Lagers der Zukunft erarbeitet. In diesem Werk nennen die Autoren ergänzend den Einsatz von Drohnen und autonomen Flurförderfahrzeugen sowie die Vorteile des 3D-Drucks und heben erneut die Relevanz der Flexibilität des Lagers hervor [Tho+17].

In der aktuellen Dekade haben sich Custodio et al. mit der Vision eines innovativen Frameworks eines flexiblen, automatisierten Lagers auseinandergesetzt. Die Forscher stellen sich dabei eine automatisierte, modulare und jederzeit rekonfigurierbare Systemstruktur innerhalb des Lagers vor. In dieser Vision wird eine RFID-basierte Steuerung implementiert und auch Custodio et al. sprechen sich gegen stationäre Strukturen aus, die sie als ein Hindernis hinsichtlich Flexibilität und Skalierung einschätzen. Aus diesem Grund nennen die Autoren die Kombination einer ausgedehnten Datenerfassung und autonomer Fahrzeuge (AGV) als die ideale Technologie für die automatisierte Einlagerung und das automatisierte Kommissionieren [CM20]. Mit einer Veröffentlichung, die das Lager der Zukunft basierend auf den von dem Unternehmen Amazon eingereichten Patenten betrachtet, bieten Delfanti et al. einen unorthodoxen Blick auf die Zukunft des Lagers [DF21]. Dabei werden vor allem Augmented Reality Lösungen betrachtet, die den Angestellten Anweisungen und Hinweise geben, automatisierte Drop-Offs durch Drohnen an sogenannten ‚Airborne Fullfilment Centers‘ sowie das automatisierte Bin Picking mittels diverser technologischer Lösungen [DF21]. Mit einem Fokus auf Mensch-Maschine-Interaktionen hat Fabian Lorson seine Dissertationsschrift „Perspectives on the Warehouse of the Future“ an der Technischen Universität München im Jahr 2023 veröffentlicht [Lor23]. Der Autor beschreibt, beispielsweise die Interaktion von Mensch und Roboter im Kontext des Order Pickings [Lor23, S. 17]. Auch in diesem Werk wird die Abwägung zwischen Automatisierung, Teilautomatisierung und manueller Arbeit thematisiert, mit dem Hinweis auf die hohen, mit dem Order Picking verbundenen Personalkosten [Lor23, S. 110]. Eine moderne, den Menschen im Lager gänzlich exkludierende Vision, ist die des ‚Dark Warehouse‘ [Mat24]. Bei dieser Vision ist selbst die Beleuchtung des Lagers obsolet, da das Lager vollautomatisiert funktioniert und somit keine Lichtquellen für menschliches Personal notwendig sind.

Zusammenfassend lässt sich festhalten, die Betrachtung des Lagers als Produkt seiner Zeit hat einen beachtlichen Wandel durchlaufen. Auch der Blick auf das Lager der Zukunft hat sich über die Jahrzehnte stark verändert. Dabei hat sich eine Entwicklung weit über die Auffassung eines Lagers als Gebäude ergeben, welches einzig und allein als Abstellplatz für Güter dient [Ack90]. Heutzutage stellt das Lager ein hochkomplexes System dar, das zunehmend erforscht und konzeptionell erweitert wird.

## 2.2 Identifikationsansätze

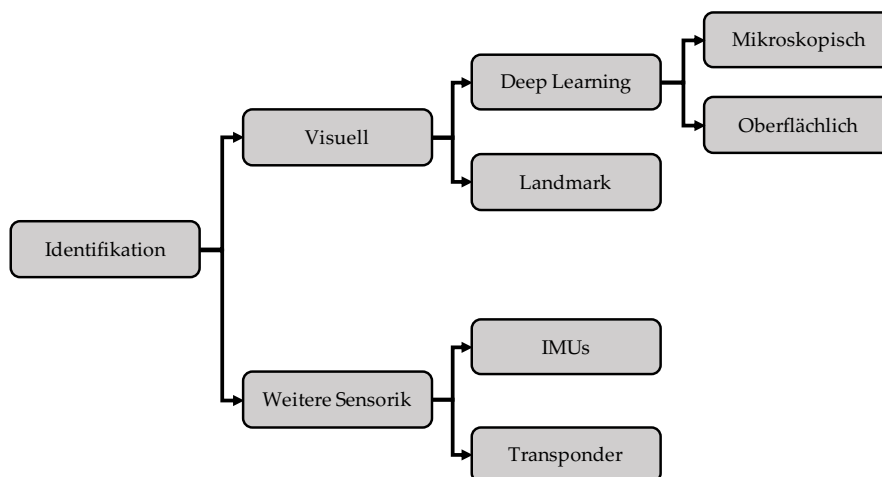
In diesem Kapitel werden gängige Identifikationsmethoden und -ansätze vorgestellt und beschrieben. Um sich näher mit diesen zu beschäftigen, ist es zunächst unerlässlich, den Begriff der Identifikation im Rahmen dieser Arbeit zu definieren.

**Definition:** *Identifikation*

In Anlehnung an bisherige Definitionsversuche in der einschlägigen Literatur [LHG12; Sag+14; Ye+21; ZYH16] wird die Identifikation im Rahmen dieser Arbeit wie folgt definiert:

Der Begriff der Identifikation beschreibt den Prozess der Zuweisung eines eindeutigen Labels (etwa in Form einer Zahlen- oder Zeichenfolge) zu einer Entität, die unter Nutzung äußerer Merkmale von anderen Entitäten derselben Art unterschieden wird. In diesem Kontext findet die Identifikation, oder Re-identifikation einer bereits aufgenommenen Entität, häufig unter Nutzung mehrerer Kameras an unterschiedlichen Standorten statt.

Wie in Kapitel 1 dargestellt beschäftigt das Konzept der Identifikation jeher Forscher diverser Fachgebiete, von der Philosophie bis zur Mathematik [LLP12]. Im Kontext dieser Arbeit liegt der Fokus dabei auf der Computer Vision basierten Identifikation von Entitäten, also abstrakten und unbelebten Gegenständen. Eine Taxonomie der Identifikation und ihrer Subkategorien ist Abb. 2.5 zu entnehmen und dient als Grundlage für die kommenden Kapitel.



**Abb. 2.5** Exemplarische Taxonomie der Identifikationsverfahren, übergreifend unterteilt in visuelle Identifikationsverfahren und Verfahren, die non-visuelle Inputs verwenden.

Die Identifikation einer Entität setzt voraus, diese sei in einer gewissen Form bereits bekannt, etwa innerhalb einer Datenbank. Nur so ist ein Abgleich und damit eine Identifikation der Entität möglich. Um diesen Prozess zu gewährleisten, wird vorangehend ebenjene Entität serialisiert. Die Serialisierung wird wie folgt definiert:

**Definition:** *Serialisierung*

Die Serialisierung (namentlich angelehnt an die Serialisierung, wie sie aus der Informatik bekannt ist [SM12]) beschreibt im Rahmen dieser Arbeit das erstmalige Einlesen einer Entität innerhalb einer Datenbank. Die Serialisierung stellt somit einen Teil der erstmaligen Identifikation dar, bei welchem das bereits beschriebene Label und die Repräsentation der Entität (in der Regel in vektorisierter Form) einer Datenbank zugeführt werden.

Vor der Serialisierung und Identifikation wird jedoch die zu identifizierende Entität zunächst visuell detektiert. Eine Definition für die visuelle Detektion wird im Folgenden gegeben:

**Definition:** *Detektion*

Der Begriff der Detektion beschreibt das Erkennen einer Entität, etwa innerhalb eines Bildes oder Videos. Die Entität wird dabei von ihrer Umgebung (den Bildinhalten, die nicht die Entität darstellen) approximativ abgegrenzt und somit segmentiert oder eingerahmt. Die Detektion setzt sich insofern von der Klassifikation ab, da bei der Klassifikation lediglich das Bild als Ganzes einer Kategorie zugeordnet wird [RW17].

Detektionsalgorithmen arbeiten gängigerweise mit annotierten Daten, die sie während des Trainings zum Abgleich mit ihren Prädiktionen nutzen. Als Ergebnis erhält der Nutzer einen Rahmen (Bounding Box) oder eine Fläche (Segmentierung), welche die relevanten Bildabschnitte kennzeichnen [Zha+19].

Im Kontrast zur Identifikation stellt die Detektion lediglich das Vorhandensein einer Entität an einer oder mehreren Stellen im Bild fest. Dies wird durch die Bewegung der Entität, abgrenzend zum sonst statischen Bild, oder durch die Ermittlung gewisser visueller Charakteristika umgesetzt [Gal+20; Mit+18; Zha+19]. Die Detektion stellt in der industriellen Praxis eine notwendige Vorstufe zur Identifikation dar, da sie zur Identifikation und Serialisierung geeignetes Bildmaterial bereitstellt. Eine graphische Gegenüberstellung der Klassifikation, Detektion und Identifikation ist Abb. 2.6 zu entnehmen.

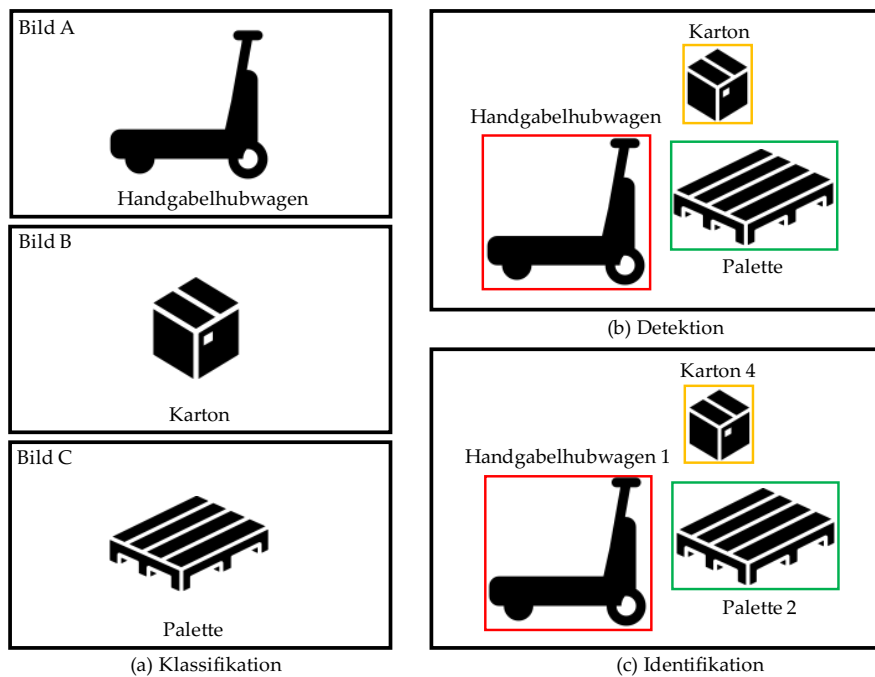


Abb. 2.6 Gegenüberstellung der Klassifikation (a), Detektion (b) und Identifikation (c).

### 2.2.1 Identifikation anhand extrinsischer Merkmale

Wie in Kapitel 1 dargelegt sind im Rahmen dieser Arbeit mit extrinsischen Merkmalen artifizielle, visuelle Eigenschaften gemeint, die eine Entität von einer anderen abgrenzen. Der Vorteil der Nutzung solcher Merkmale ist die mehrheitlich einfache Einlesbarkeit und Kontrolle über den Scan- oder Leseprozess.



Abb. 2.7 Beispiele extrinsischer Kodierungen: DataMatrix (a), UPC-A (b) und QR (c).

Ein weit verbreitetes Beispiel für extrinsische Merkmale sind visuelle Codierungen (siehe Abb. 2.7), wie der Barcode. Barcodes sind optische Dar-

stellungen von Daten, die zur automatischen Identifizierung von Objekten verwendet werden [DK05]. Patentiert im Jahr 1952 [WB52] bestehen sie klassischerweise aus einer Reihe von parallelen Balken. In ihrer eindimensionalen Form (1D-Codes oder auch lineare Codes genannt) repräsentieren die Balken und die Abstände zwischen ihnen unterschiedliche Informationen. Industrielle Verwendung findet der Barcode beispielsweise unter Nutzung der UPC (Universal Product Code) oder EAN (European Article Number) Symbologie, genormt nach ISO/IEC 15420 [ISO/IEC 15420:2009]. In Abb. 2.7 wird exemplarisch das UPC-A Format verwendet. Der Barcode ist auch in der Logistikforschung vertreten [Gou13; Hon09], beispielsweise um Wareneingänge und -ausgänge zu verfolgen. Beim Scannen eines Barcodes wird der Code in eine digitale Form umgewandelt und die darin enthaltenen Daten anschließend verarbeitet. Durch das Scannen eines Barcodes mit einem Barcode-Scanner oder einem Smartphone werden Daten schnell erfasst und in Computersysteme oder Datenbanken eingelesen.

Barcodes werden in vielen Branchen eingesetzt, darunter der Einzelhandel, das Gesundheitswesen und die Lagerverwaltung [Ist+20]. Sie ermöglichen in diesen Branchen eine schnelle und genaue Identifikation von Produkten, das Verfolgen von Beständen, das Erfassen von Informationen über Produkteigenschaften. Die Verwendung von Barcodes bietet mehrere Vorteile, wie zum Beispiel eine Effizienzsteigerung, Fehlerreduzierung, automatisierte Datenerfassung, verbesserte Bestandsverwaltung und beschleunigte Transaktionsabwicklung. Menschliches Versagen beim Ablesen und Übertragen einer Nummer sind somit ausgeschlossen. Sie haben sich als kostengünstige und zuverlässige Methode zur Erfassung und Verwaltung von Daten etabliert und sind ein essentieller Bestandteil der Automatisierung und Verbesserung von Prozessen in diversen Branchen.

Während lineare Barcodes aus einer Reihe von parallelen Linien unterschiedlicher Breite bestehen, die lediglich entlang einer Achse über Informationen verfügen, codieren zweidimensionale Barcodes Informationen sowohl horizontal als auch vertikal. Auf diese Weise speichern zweidimensionale Barcodes eine größere Datenmenge. Beispiele für zweidimensionale Barcodes sind QR-Codes und DataMatrix-Codes, wie sie in Abb. 2.7 zu sehen sind.

DataMatrix-Codes sind zweidimensionale Barcodes, die Informationen in Form von schwarz-weißen Quadraten speichern. Die Codierung von DataMatrix-Codes erfolgt durch die Platzierung von Quadraten in einem Raster. Die Daten werden binär codiert, wobei die schwarzen Quadrate als Einsen und die weißen Quadrate als Nullen interpretiert werden. Dank Fehlerkorrekturmechanismen ist der Code auch bei beschädigten oder teilweise unlesbaren Quadraten weiterhin lesbar. Im Gegensatz zu linearen Barcodes sind DataMatrix-Codes in beliebigen Orientierungen lesbar, da die Informationen sowohl horizontal als auch vertikal codiert sind [GS1 DMG; Pla95].

DataMatrix-Codes werden in verschiedenen Anwendungen eingesetzt, darunter die Produktkennzeichnung, Dokumentenverwaltung und medizinische Verpackungen [CP19; KP20; Zha+18]. Sie ermöglichen die eindeutige Identifikation von Produkten, die Verfolgung von Beständen und die automatische Datenerfassung. Durch das Scannen des DataMatrix-Codes werden Informationen wie Seriennummern, Chargennummern, Produktionsdaten oder andere benutzerdefinierte Daten abgerufen.

DataMatrix-Codes bieten einige Vorteile gegenüber linearen Barcodes, wie zum Beispiel die Möglichkeit, eine größere Datenmenge zu speichern, die Robustheit gegenüber Beschädigungen und die flexiblere Platzierung auf unterschiedlichen Oberflächen, bei Erhalt der Lesbarkeit. Sie werden von verschiedenen Scangeräten, wie Handheld-Scannern oder Smartphones, gelesen, die über die entsprechende Dekodierungssoftware verfügen.

Wie DataMatrix-Codes sind auch QR-Codes (Quick Response Codes) zweidimensionale Barcodes, die Informationen in Form von schwarz-weißen Mustern speichern [Soo08]. Sie wurden entwickelt, um schnell und effizient Informationen zu erfassen und zu übertragen und bestehen ebenfalls aus einem Raster von schwarzen und weißen Quadraten, die Informationen in binärer Form repräsentieren [Tiw16]. Die Codierung enthält verschiedene Arten von Daten. Üblich sind dabei kurze Texte, URLs, Kontaktdaten oder multimediale Inhalte. QR-Codes werden via Dekodierungssoftware auf Smartphones oder anderen Endgeräten ausgelesen, um die in ihnen enthaltenen Informationen zu lesen und anzuzeigen [Soo08]. QR-Codes werden häufig in Marketingkampagnen eingesetzt, um etwa einem Kunden zusätzliche Informationen oder spezielle Angebote bereitzustellen. Durch das Scannen des QR-Codes mit einem Smartphone gelangen Benutzer beispielsweise auf eine Website mit weiteren Produktinformationen und bieten Unternehmen damit eine Möglichkeit, ihre Zielgruppe zu erreichen und interaktive Erlebnisse zu schaffen [Tiw16].

Der QR-Code wurde in den 1990er Jahren im Kontext der japanischen Logistik entwickelt [Har19] und wird auch heutzutage vielfältig in der Logistik eingesetzt, um Prozesse effizienter zu gestalten. QR-Codes werden an Produkten, Paletten oder Behältern angebracht, um den Bestand zu verfolgen und zu verwalten. Durch das Scannen des QR-Codes haben Mitarbeiter die Möglichkeit schnell auf Informationen zu Lagerort, Bestandsmenge und Produktmerkmalen zuzugreifen. Dies ermöglicht eine effiziente Inventur, Nachbestellung und Verfolgung von Warenströmen. QR-Codes werden auf Versandetiketten angebracht, um eine eindeutige Identifizierung und Verfolgung von Sendungen zu ermöglichen. Transportunternehmen scannen QR-Codes, um den aktuellen Standort und den Lieferstatus einer Sendung zu erfassen [Tiw16]. Dies ermöglicht eine Echtzeitüberwachung und verbesserte Transparenz entlang der Lieferkette. Weiter werden QR-Codes auf Fahrzeugen oder Transportbehältern angebracht, um eine einfache Identifikation und

Zuordnung dieser zu ermöglichen. Durch das Scannen des QR-Codes sind Fahrzeugdaten, Wartungshistorie oder spezifische Ladevorschriften abrufbar [MI15]. Dies hilft bei der Verwaltung von Flotten, der Zuweisung von Ressourcen und der Optimierung des Fahrzeugbetriebs. QR-Codes werden in der Logistik verwendet, um Informationen zu Sicherheitsvorschriften, Qualitätsstandards oder Produktspezifikationen bereitzustellen. Mitarbeiter greifen durch das Scannen des QR-Codes auf relevante Dokumente, Anweisungen oder Schulungsmaterialien zu. Dies gewährleistet die Einhaltung von Standards und verbessert die Qualitätssicherung. Des Weiteren sind QR-Codes bei der Verarbeitung von Rücksendungen oder Retouren einsetzbar. Durch das Scannen des QR-Codes auf einem Rücksendetikett werden Informationen zur ursprünglichen Bestellung, dem Grund der Rücksendung und dem weiteren Verarbeitungsprozess abgerufen. Dies erleichtert eine effiziente Bearbeitung und Rückverfolgung von Retouren. Die Anwendungsmöglichkeiten von QR-Codes in der Logistik sind vielfältig und tragen zur Verbesserung der Prozesseffizienz, Transparenz und Bestandsverwaltung bei. Durch die Integration von QR-Codes in die logistischen Abläufe wird es Unternehmen ermöglicht, ihre Kosten zu senken, Fehler zu reduzieren und eine bessere Kundenbetreuung zu gewährleisten. Allerdings besteht auch ein potenzielles Risiko für Missbrauch und Betrug. So ist es möglich, QR-Codes mit Links zu betrügerischen Websites, Malware oder Phishing-Attacken zu versehen [Kro+14].

Über codebasierte Lösungen hinaus, ist weiterhin RFID (Radio Frequency Identification) [Wei07] als eine weit verbreitete Technologie zu nennen, die es ermöglicht, Objekte zu identifizieren und zu verfolgen. RFID-Lösungen haben in den letzten Jahren zunehmend an Bedeutung gewonnen und werden in verschiedenen Anwendungsbereichen wie Logistik, Einzelhandel, Gesundheitswesen und Transport eingesetzt [CBB22]. RFID besteht aus drei Hauptkomponenten: einem (aktiven oder passiven) RFID-Tag, einem Lesegerät und einem Computersystem [Wei05]. Der RFID-Tag enthält einen Mikrochip und eine Antenne, die es dem Tag ermöglichen, über Radiowellen mit dem Lesegerät zu kommunizieren. Das Lesegerät sendet elektromagnetische Signale aus, um den RFID-Tag zu aktivieren und Informationen abzurufen. Die gesammelten Daten werden dann an das Computersystem übertragen, wo sie verarbeitet und weiterverwendet werden. Im Einzelhandel wird RFID beispielsweise zur Verwaltung von Warenbeständen und für die Diebstahlsicherung eingesetzt. RFID-Tags ermöglichen eine schnelle Inventur, automatische Nachbestellungen und die Verfolgung von Produkten von der Produktion bis zum Verkaufsort. RFID-Lösungen werden häufig eingesetzt und bieten viele Vorteile gegenüber codebasierten Lösungen. Dennoch wird davon ausgegangen, RFID stelle stets die kostspieligere Lösung dar und ist somit nicht in jedem Anwendungsfall ein geeigneter Ersatz für eine codebasierte Lösung [AT12].

### 2.2.2 Identifikation anhand inhärenter Merkmale

Einerseits bieten standardisierte, extrinsische Methoden wie Barcodes oder QR-Codes zuverlässige Serialisierungsansätze, die bereits weitläufig etabliert sind. Demgegenüber steht die Notwendigkeit, die relevanten Entitäten mit artifiziellen Identifikatoren (Codes oder Transpondern) auszustatten, womit zusätzliche Arbeitsschritte und Kosten verbunden sind. Zusätzlich hängt die Funktionalität der Serialisierung vollständig von der Unversehrtheit der genutzten Codes oder Transponder ab, die in industriellen Umgebungen nicht zwangsläufig gewährleistet ist.

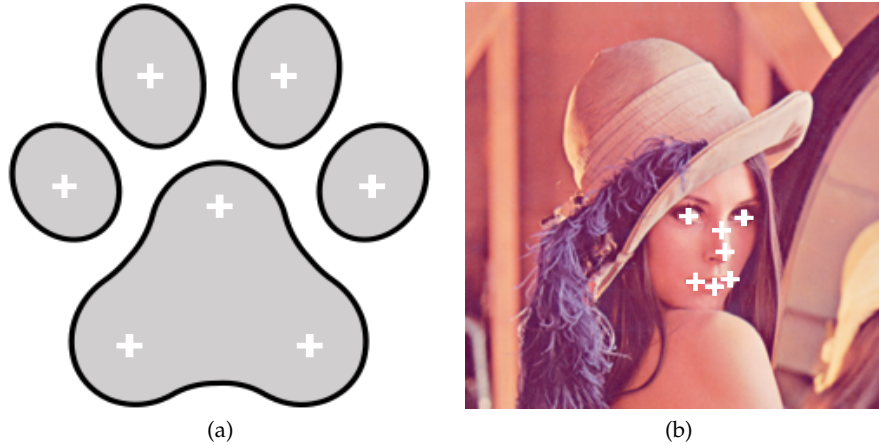
Vor diesem Hintergrund stellt sich die Frage, ob Computer Vision Methoden aus dem Bereich der Identifikation von Personen, auf logistische Entitäten übertragbar sind. In bereits existierenden Forschungsarbeiten wurden mehrfach Methoden zur Identifikation von Materialien anhand ihrer inhärenten Eigenschaften untersucht. So ist die FIBAR (Fingerprint Imaging by Binary Angular Reflection) Methode zu nennen, die mikroskopische Aufnahmen der Oberflächenstruktur metallischer Gegenstände nutzt, um diese voneinander zu unterscheiden [TKI17]. Diese Methode beschränkt sich allerdings auf die Nutzung metallischer, einheitlich hergestellter Gegenstände und benötigt mikroskopische Aufnahmen ihrer Oberflächenstruktur.

Ein ähnlicher Ansatz wurde in einer Dissertationsschrift der Universität Stuttgart und ihren Vorarbeiten beschrieben [Wig+18; Wig20]. Der darin dokumentierte Ansatz verfolgt die Nutzung von Nahaufnahmen von Molded Interconnect Devices (MIDs) mit monochromer Domebeleuchtung. Es werden dabei spezifisch Ausschnitte mit Flächen von bis zu  $37 \text{ mm} \times 37 \text{ mm}$  untersucht und selbst leichte Beleuchtungsschatten führen bereits zu „[...] deutlichen Verschlechterungen der Identifikationssicherheit [...]“ [Wig20]. Dieser Ansatz ähnelt insofern der FIBAR Methode, da seine industrielle Anwendung und Generalisierbarkeit durch die genannten Limitierungen fraglich ist.

Mit dem Fokus auf den Werkstoff Holz wurde bereits im Jahr 1995 die Dissertationsschrift von Anton Klar an der Technischen Universität Dortmund veröffentlicht [Kla95]. Hierbei wurden klassische Bildverarbeitungsverfahren verwendet, um Ladehilfsmittel anhand ihrer stochastischen Oberflächeninformationen voneinander zu unterscheiden. Die Datentransparenz dieser Arbeit ist aufgrund ihres Alters allerdings eingeschränkter Natur. Der technische Ansatz unterscheidet sich außerdem durch die Nutzung von ‚Hand-Crafted Features‘ von modernen Computer Vision Verfahren. Die Nutzung solcher Features beschränkt die Generalisierbarkeit des Verfahrens weiter. Darüber hinaus existieren weitere Ansätze zur reinen Unterscheidung diverser Holzarten [HW11].

Im Allgemeinen sind klassische Computer Vision Lösungen, die im Vorfeld üblicherweise sogenannte ‚Landmarks‘ oder die bereits genannten Hand-

Crafted Features definieren, optimiert für einen speziellen Anwendungsfall. So werden beispielsweise die Position und Abstände von Augen, Nase und Mund als Landmarks für die Identifikation einer Person anhand ihrer Gesichtsm Merkmale definiert [Kor+20] (siehe Abb. 2.8).



**Abb. 2.8** Beispiele der Nutzung von Landmarks, dargestellt als weiße Kreuze, für die Identifikation von Tieren (a) und Menschen (b). Exemplarisch genutzt sind hierfür ein Pfotenabdruck und das ‚Lena Testbild‘ für Bildverarbeitung von Dwight Hooker (1972).

Eine weitere Anwendung wäre die Identifikation von spezifischen Tieren innerhalb einer Gruppe [Jew+16; Li+18; Lok+14], wobei die markantesten Merkmale der jeweiligen Pfotenabdrücke als Landmarks genutzt werden. Auch bei diesen Anwendungsfällen entstehen Datenbanken, die je nach Anwendung das Gesicht einer Person oder den Pfotenabdruck eines Tiers lokal speichern. Das bedeutet, eine de facto Serialisierung (siehe Kapitel 2.1) des betreffenden Subjekts hat anhand seiner inhärenten visuellen Merkmale stattgefunden. Dennoch sind die Ansätze nicht austauschbar nutzbar, weshalb die Landmarks für menschliche Gesichter nicht für Pfotenabdrücke nutzbar sind und vice versa.

Ein Forschungsbereich, der ohne die Nutzung festgelegter Landmarks auskommt, ist die Identifikation von Personen. Häufig eingesetzt in der Überwachung des öffentlichen Raums bedienen sich diese Computer Vision Verfahren aber ebenfalls an Regelmäßigkeiten, wie dem aufrechten Gang eines Menschen, um Bildsegmentierungen für eine zuverlässigere Identifikation zu gewährleisten [Sun+18]. Die eigentliche Identifikation ist dabei oft Teil eines Triumvirat aus einer initialen Detektion, gefolgt von Tracking und letztlich Identifikation [ZYH16]. Der Forschungsbereich der Identifikation von Personen wächst seit seinen Anfängen um den Millenniumswechsel herum [HR97; ZZK05] stark an, besonders dank der Nutzung und Entwicklung von Deep Learning Methoden [Wu+19; ZYH16]. Für den Anwendungsfall

der Identifikation von Personen werden entweder einzelne Bilder oder Videos (Bildsequenzen) genutzt [ZYH16]. Es werden in der Regel keine Landmarks verwendet, jedoch stellenweise initiale Featuregewichtungen [ZYH16]. Unabhängig von der Gewichtung und Erstellung der Features werden die Repräsentationen einer Person in aller Regel in Form von Feature-Vektoren in einer Datenbank gespeichert.

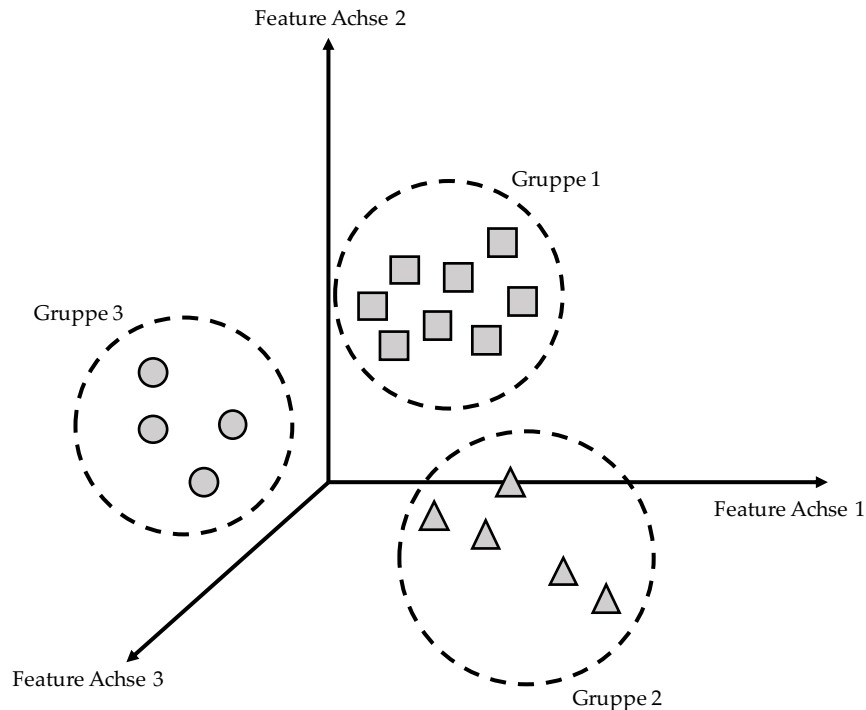


Abb. 2.9 Exemplarische Darstellung eines Embedding Space.

Bei einem gelungenen Trainingsprozess repräsentiert dieser Vektor die respektive Person möglichst präzise und grenzt sie damit von anderen Personen in der Datenbank ab. Um zwei Feature-Vektoren miteinander zu vergleichen, werden üblicherweise Distanzen zwischen ihnen berechnet. Diese Distanz ist möglichst gering, wenn es sich um die gleiche Person handelt und möglichst groß, wenn es sich um verschiedene Personen handelt. Es bilden sich somit in einem Embedding Space (auch Latent Space genannt) diverse Cluster, welche die Positionen der Feature-Vektoren der einzelnen Individuen möglichst gruppiert darstellen (siehe Abb. 2.9). Dieser Prozess wird als Metrisches Lernen (Deep Metric Learning) bezeichnet [Wu+19; ZYH16]. In der aktuellen Forschung haben sich Standards für die Durchführung dieser Lernverfahren etabliert. Ein Beispiel ist die Nutzung der Triplet Loss und

vormals Contrastive Loss Funktionen [Min+22; Wu+19; Ye+21]. Die Triplet Loss Funktion bestimmt sich wie folgt:

$$L_{\text{triplet}}(A, P, N) = \max\{d(A, P) - d(A, N) + \alpha, 0\} \quad (2.2)$$

Verwendet wird bei dieser Loss Funktion ein Ankerbild (das Bild, das es zu vergleichen gilt)  $A$ , ein positives (ähnliches)  $P$  und ein negatives (unähnliches) Beispielbild  $N$ , deren Distanz  $d$  ermittelt wird [Min+22; Wu+19; Ye+21]. Die Verwendung der Triplet Loss Funktion hat sich im Bereich der Identifikation von Personen etabliert [HBL17]. Begründet ist dies unter anderem an der Fähigkeit der Triplet Loss Funktion, das Phänomen des Dimensional Collapse zu minimieren [Lev+21]. Hierbei handelt es sich um ein Phänomen, bei dem Repräsentationen in niedrigdimensionale Embedding Spaces komprimiert werden, wodurch wichtige Unterscheidungsmerkmale zwischen Identitäten oder Klassen verloren gehen. Im Kontrast dazu, zielt die Triplet Loss Funktion darauf ab, ähnliche Identitäten im Embedding Space näher beieinander zu positionieren und unterschiedliche Identitäten weiter voneinander zu entfernen. Die Distanzen zwischen Identitäten im Embedding Space werden auf unterschiedliche Weisen berechnet, unter anderem mittels Euklidischer oder Cosinus Distanz [WB18].

Eine Alternative zur Vektorrepräsentation stellt die Graphrepräsentation dar. Als Graph wird hier beschrieben als  $G = (V, E)$  wobei  $V$  eine Menge von Knoten und  $E$  eine Menge von Kanten repräsentiert. Gängigerweise werden Graphen auch als Knowledge Graphs genutzt, wobei ihre Knoten dabei Konzepte und ihre Kanten (abstrakte) Beziehungen zwischen den jeweiligen Konzepten repräsentieren [Rue+21]. Die Größe (Gewichtung) der Knoten und Dimensionalität (Länge) der Kanten ist ungewichtet oder gewichtet darstellbar [Rue+21]. Graphen werden im Kontext des Deep Learning auch als Graph Neural Networks (GNNs) verwendet. Im Fall von Bilddaten werden Convolutional GNNs verwendet, die Beziehungen und Strukturen in unregelmäßigen, strukturierten Daten modellieren, während sich Deep Neural Networks (DNNs) auf die Verarbeitung von Daten mit fester Rasterstruktur (innerhalb quadratischer Pixelpools) konzentrieren [Wu+20]. Die Nutzung von Graphen reicht von der Medizin [Ago+13], hin zur Partikelphysik [Fey49] und zur Linguistik [Pip14]. Die Nutzung von GNNs im Maschinellen Lernen hat in diesem Zuge an Bedeutung gewonnen [Zho+20], so etwa in der bildlichen Darstellung von Informationen. Analog zur Vektorrepräsentation bietet die Nutzung von GNNs vielversprechende Ergebnisse [Ave+20]. Auch für die Identifikation von Personen wurden GNNs bereits eingesetzt [Che+18a; She+18] und dienen dazu, die Triplet Loss Funktion effizient zu minimieren.

Des Weiteren wird bei der Betrachtung des Identifikationsproblems eine Unterscheidung zwischen der ‚Open World‘ und ‚Closed World‘ Annahme

unternommen [ZYH16]. Bei einem Closed World Ansatz wird davon ausgegangen, jede zugeordnete Person sei bereits Teil des Datensatzes. Bei einem Open World Ansatz wird ein Neuheitsschwellwert (Novelty Threshold) definiert, bei dessen Überschreitung von einer neuen Person ausgegangen wird, die es dem Datensatz hinzuzufügen gilt [Ye+21]. Eine Herausforderung stellt bei der Identifikation von Personen die Annotation dar, weshalb auch das Unsupervised Learning in diesem Bereich erforscht wird [Ye+21]. Je nach Benchmarkdatensatz erreichen state-of-the-art Modelle aktuell Genauigkeiten zwischen 81 % und 97 % [Min+22].

Zur Quantifikation der Leistung eines closed-world Identifikationsverfahrens wird ein sogenanntes Query Bild mit der Gallery verglichen. Es handelt sich bei einem Query Bild um ein Testbild, das eine ID beinhaltet, die in der Gallery (der Datenbank bekannter IDs) ebenfalls beinhaltet ist. Das Bild der Gallery, das die größte Ähnlichkeit zum Query Bild aufweist, wird als erste Zuordnung bestimmt. Zur konkreten Evaluation werden unterschiedliche Metriken verwendet. Zunächst ist die Accuracy  $A$  (Genauigkeit) zu nennen. Diese berechnet sich wie folgt aus den Summen der falschen und richtigen Zuordnungen (True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)) der Identitäten eines Datensatzes:

$$A = \frac{TN + TP}{TN + FN + TP + FP} \quad (2.3)$$

Die Accuracy wird darüber hinaus in einem Ranglistenverfahren ausgewertet (Ranked Accuracy oder auch Rank-k-Accuracy) [Min+22; Wu+19; Ye+21; ZYH16]. Das Konzept der Ranked Accuracy ist im Rahmen dieser Arbeit wie folgt zu verstehen:

**Definition:** *Ranked Accuracy*

Unter Annahme einer Liste von Testfällen, wobei jeder Testfall aus einem Paar von tatsächlicher Identität  $y_i$  (Ground Truth) und den vorhergesagten Wahrscheinlichkeiten  $p_{i_1}, p_{i_2}, \dots, p_{i_n}$  für jede der  $n$  möglichen Identitäten besteht, wird die Accuracy für einen bestimmten Rang  $k$  (Rank-k-Accuracy) bestimmt als  $\frac{1}{N} \sum Match(y_i, k)$ . Hierbei ist  $N$  die Gesamtanzahl der Testfälle und  $Match(y_i, k)$  eine Funktion, die den Wert 1 ausgibt, wenn die tatsächliche Identität  $y_i$  auf einer Position  $\leq k$  in den vorhergesagten Wahrscheinlichkeiten erscheint. Andernfalls gibt die Funktion den Wert 0 aus.

Algorithmisch ist die Ranked Accuracy darüber hinaus wie folgt formuliert:

**Algorithmus 1** Ranked Accuracy Algorithmus.

---

**Input:** Prädiktionen  $P$ , Ground Truth  $GT$ , Rang  $k$   
 $num\_correct \leftarrow 0$  ▷ Anzahl korrekt zugeordneter Identitäten  
 $total\_samples \leftarrow \text{length}(GT)$  ▷ Gesamtanzahl Identitäten  
**for**  $i \leftarrow 1$  to  $total\_samples$  **do**  
   $sort\ indices$   
  **if**  $GT[i] \in \text{sorted\_indices}[1 : k]$  **then**  
     $num\_correct \leftarrow num\_correct + 1$   
  **end if**  
**end for**  
**Output:** Rank- $k$ -Accuracy:  $\frac{num\_correct}{total\_samples}$

---

Aus den bereits für die Accuracy  $A$  verwendeten Maßen werden wiederum Precision  $P$  (Präzision) und Recall  $R$  (Sensitivität) bestimmt:

$$P = \frac{TP}{TP + TN} \quad (2.4)$$

$$R = \frac{TP}{TP + FN} \quad (2.5)$$

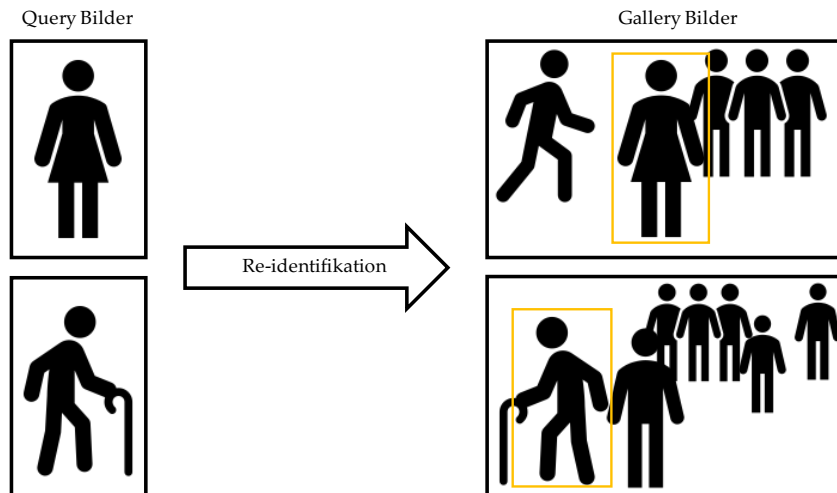
Des Weiteren wird aus der Precision die mean Average Precision (mAP) bestimmt:

$$mAP = \frac{1}{n} \sum AP \quad (2.6)$$

Um die Konfidenz der erhaltenen Ergebnisse zu erhöhen und ihre Validität über ein spezifisches Subset der Gesamtdaten hinaus zu garantieren, ist es ratsam eine Kreuzvalidierung durchzuführen [Sch93; VB12].

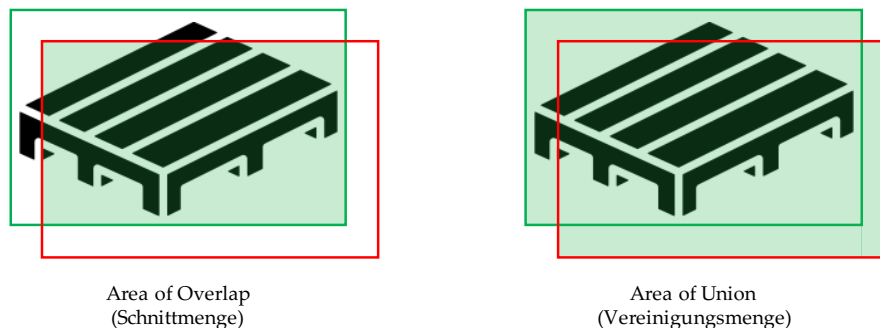
Bevor jedoch die eigentliche Identifikation stattfindet, gilt es, die zu identifizierende Entität oder das zu identifizierende Subjekt im Bild zu detektieren [ZYH16]. Dazu werden Object Detection Modelle vorgeschaltet, die relevante Entitäten aus dem Bild extrahieren, beispielsweise YOLO (You Only Look Once) [Red+16]. YOLO ist ein Modell, das Objekte innerhalb von Bildern oder Videosegmenten (Standbildern) erkennt, indem es das Bild in ein Raster unterteilt und Bounding Boxes sowie, im Falle mehrerer Klassen, Klassenwahrscheinlichkeiten vorhersagt. Im Gegensatz zu seinen Vorgängern erfordert YOLO keine mehrfachen Durchläufe durch das Netzwerk und verarbeitet stattdessen das gesamte Bild in einem einzigen Forward Pass, was es schneller und effizienter macht. Aufgrund dessen wird YOLO nach wie vor weit verbreitet genutzt und wurde seit seiner Einführung weiter angepasst und adaptiert. Eine Adaption von YOLO stellt Tiny YOLO dar. Dieses Modell verfügt über 9 anstatt 24 Convolutional Layers, die jeweils

niedrigere Filteranzahlen aufweisen. Basierend auf der Bounding Box, mit welcher Object Detection Modelle wie YOLO die Position der Entität im Bild zu approximieren versuchen (siehe Abb. 2.10), wird der letztendliche Input für das Identifikationsmodell erstellt.



**Abb. 2.10** Beispielhafte Veranschaulichung des Vorgehens bei der Identifikation von Personen. Dargestellt ist der Abgleich von Query und Gallery sowie die Detektion von Personen, visualisiert als gelbe Bounding Box.

Sofern annotierte Trainingsdaten vorhanden sind (Labeling und korrekte Bildposition der Entität), wird die Position der Bounding Box, die vom Modell bestimmt wurde, mit der Ground Truth verglichen (siehe Abb. 2.11).



**Abb. 2.11** Exemplarische Darstellung der Komponenten der Intersection over Union: Links Area of Overlap, rechts Area of Union. In Grün dargestellt ist die Ground Truth, in Rot die Prädiktion.

Neben der Position sind auch Größe und Form der Bounding Box für eine akkurate Detektion relevant. Das Ziel ist es somit, den Wert der ,Inter-

section over Union' ( $IoU$ ) zu maximieren. Die  $IoU$  berechnet sich aus der Schnittmenge der vorhergesagten Fläche  $P$  und der Ground Truth  $GT$  (Area of Overlap ( $A_o$ )) sowie ihrer Vereinigungsmenge (Area of Union ( $A_u$ )):

$$IoU = \frac{A_o}{A_u} = \frac{|P \cap GT|}{|P \cup GT|} \quad (2.7)$$

## 2.3 Datensatzerstellung im Maschinellen Lernen

Für das erfolgreiche Training eines datengetriebenen Deep Learning Modells ist das Vorhandensein eines geeigneten Datensatzes nötig. In vielen Fällen werden bereits vorhandene Datensätze verwendet, die sich weitestgehend als Benchmark etabliert haben. Beispiele dafür sind MNIST [LeC+98], Fashion-MNIST [XRV17] und ImageNet [Den+09] im Bereich der Klassifikation, COCO [Lin+14] und KITTI [Gei+13] im Bereich der Detektion und Market-1501 [Zhe+15], CUHK03 [Li+14] und VIPeR [GBT07] im Bereich der Re-identifikation. Diese Datensätze dienen datengetriebenen Ansätzen als ausschlaggebende Grundlage. Von hoher Relevanz ist dabei eine sinnvolle Auswahl an Sensoren und der Art, wie aus ihnen Informationen in maschinenlesbare Formate konvertiert werden (siehe Kapitel 2.1.2).

Auch wenn Modelle auf gängigen Datensätzen vortrainiert werden [Erh+10], um anschließend auf nicht im Datensatz vorhandene Subjekte angewendet zu werden, ist in einem solchen Fall die Anpassung des Modells für den spezifischen Anwendungsfall notwendig. Ein solcher Fall tritt immer dann auf, wenn es sich um einen besonderen Anwendungsbezug handelt. Ein Beispiel dafür ist eine Nische, für die keine Datengrundlage besteht. In einem solchen Fall ist es wichtig, einen hochwertigen und für den Anwendungsfall sinnvollen Datensatz zu erstellen.

Dennoch mangelt es aktuell an Best Practices und Verfahrensvorgaben hinsichtlich der Erstellung von Datensätzen [Hut+21; SHD21]. Dies ist vermutlich aufgrund der Dynamik und des noch jungen Alters des modernen Forschungsfeldes des datengetriebenen Maschinellen Lernens der Fall. Einige Forscher zweifeln aus diesem Grund gar an der Reproduzierbarkeit der Ergebnisse, die in diesem Forschungsbereich publiziert werden [Hut18]. Erste Versuche hinsichtlich einer vereinheitlichten Vorgehensweise oder einer Nutzung von Best Practices werden jedoch unternommen [Rei+23].

Über die Reproduzierbarkeit der Ergebnisse hinaus, stehen oftmals weitere Ansprüche an Datensätze im Widerspruch zueinander oder sind Ausschlusskriterien für die Nutzbarkeit in einem spezifischen Anwendungsfall. Vereinheitlichte Vorgehensweisen und hohe Ansprüche an die Datenqualität etwa, stehen im Kontrast zu realistischen Aufnahmen, wenn es um die Anwendung im industriellen Umfeld geht. Weiter ist es möglich eine Vielzahl an Ansprüchen an einen Datensatz zu erfüllen, während dieser dennoch nicht nutzbar für einen spezifischen Anwendungsbereich ist. So divergieren bei-

spielsweise Klassifikations- und Detektionsdatensätze in ihren Ansprüchen an die vorhandenen Daten, wie in Abb. 2.6 dargestellt wurde.

## 2.4 Verlässlichkeit als Aspekt des Maschinellen Lernens

Der Begriff der Verlässlichkeit wird in der einschlägigen Literatur aktuell mehrdeutig verwendet, ist facettenreich und auch von Interesse für diverse Forschungsgebiete außerhalb der Naturwissenschaften. Zu nennen sind dabei die Philosophie, Psychologie und Ökonomie [AS21, S. 71].

Im Kontext des Maschinellen Lernens werden anhand des Begriffs unterschiedliche Aspekte eines Verfahrens betrachtet. Es handelt sich dabei um die Datensätze und ihre Erfassung, Verarbeitung, Sicherung und etwaige Zugriffsrechte, aber auch um die Auswahl, das Training oder die Anpassung eines Modells. Die durch das Modell erzielten Ergebnisse und deren Auswirkungen, in Abhängigkeit des Anwendungsfelds, stellen einen weiteren Aspekt der Verlässlichkeit dar. All diese Aspekte sind letztendlich hinsichtlich ihrer Verlässlichkeit überprüfbar, sofern letztere definiert und quantifizierbar ist.

Der Datenschutz ist in diesem Kontext von Interesse, besonders hinsichtlich der Datenerfassung [GP08; Lie15]. Dieser ist auch im rechtlich-regulatorischen Rahmen ein relevantes Thema, etwa im Kontext der Datenschutzgrundverordnung (DSGVO) der Europäischen Union. Über den reinen Schutz von personenbezogenen Daten hinaus gilt es auch mögliche Verzerrungen in Datenstrukturen zu detektieren. Gemeint ist damit der Fehler eines Modells und seine einhergehende Abweichung von der Ground Truth. Methoden, die ebenjene Verzerrungen zu minimieren versuchen, verfügen über eigene Quantifikationsmetriken, beispielsweise den Propensity Estimator [JSS17]. Darüber hinaus ist die jeweils verwendete Rechnerarchitektur sicherheitsrelevant und zu berücksichtigen, insbesondere im Hinblick auf die vertrauenswürdige Ausführung von Modellen [Pri20]. Umfassende Untersuchungen des Stands der Technik hinsichtlich des Sicherheitsaspekts im Maschinellen Lernen werden beispielsweise von Houben et al. [Hou+22b] und Huang et al. [Hua+20] gegeben, insbesondere hinsichtlich Herangehensweisen zur Verifikation von Modellfunktionalitäten mittels Adversarial Attacks.

Huang et al. thematisieren dabei auch die Interpretierbarkeit und Erklärbarkeit als wichtigen Faktor im Kontext des Maschinellen Lernens. Gemeint ist damit die Möglichkeit für einen Nutzer, das Handeln eines Modells nachzuvollziehen [Hua+20]. Da es aktuell nicht gänzlich möglich ist Modelle zu erstellen, die frei von jeglichem Bias funktionieren, gewinnt die Interpretierbarkeit der Modelle an Relevanz. Einige Modellarten, wie Entscheidungsbäume und Support Vector Machines (SVMs), sind transparenter als

Deep Learning Modelle und ermöglichen den Nutzern ein ausgeprägteres Verständnis des Modells [Rüp06]. Darüber hinaus helfen Qualitätsmaße und Kausalitätsanalysen den Nutzern, die Ergebnisse von Modellen zu interpretieren [Hol+20]. Die Angabe der Unsicherheit von Modellergebnissen (beispielsweise einer Prädiktion) stellt einen weiteren Schritt hinsichtlich einer besseren Interpretation von Modellen dar. So wird die Interpretierbarkeit von DNNs durch die Einführung des Monte-Carlo Dropout oder der Monte-Carlo Batch Normalization erhöht [TAS18].



**Abb. 2.12** Visualisierung der entscheidenden Features eines Klassifikationsmodells als Attention Heatmap (basierend auf dem letzten Convolutional Layer des Modells ResNet50, visualisiert mittels Grad-CAM [Sel+17]).

Die Erklärbarkeit von DNNs bleibt jedoch eine besondere Herausforderung im Kontrast zu anderen (shallow) Modellen. Mehrere Arbeiten stellen in diesem Kontext Visualisierungen der einzelnen Layers und der von DNN gelernten Features als Attention Heatmap [SWM17; Tjo+21] oder Peak Response Maps (PRMs) vor [Zho+18], siehe Abb. 2.12. Besonders für Aufgaben der Bildklassifikation stellen die schichtweise Relevanzfortpflanzung [Mon+19] oder die Gewichtung von Aktivierungsunterschieden [SGK17] alternative Ansätze dar. Auch eine Angabe der Unsicherheit von Prädiktionen unterstützt eine bessere Interpretation von Modellen, wie es bei anderen Maschinellen Lernmethoden ein gängiger Ansatz ist [HPM18]. Zuletzt stellen bezüglich des Aspekts der Erklärbarkeit Chazette et al. und Zhou et al. die Frage, ob Anwender explizite Erklärungen zu den Ergebnissen der verwendeten Modelle erhalten möchten [CKS19] und wie die Angabe von Entscheidungseinflüssen das Vertrauen der Anwender an die Modelle beeinflusst [Zho+19a]. Dementsprechend ist die Interpretierbarkeit gegebenenfalls

an die respektiven Anwender anzupassen, wodurch Unterschiede in der Visualisierung und dem Detailgrad der Aussagegüte impliziert sind.

Der Begriff der Fairness wird ebenfalls als ein Aspekt der Vertrauenswürdigkeit verstanden [PS22]. Um ein faires Modelldesign zu ermöglichen, gilt es Systeme des sozialen Vertrauens zu implementieren, welche die Integrität der Beteiligten in diesem System garantieren, etwa die respektiven Entwickler und Behörden [AS21, S. 71-72]. Entgegen anderer Aspekte der Verlässlichkeit, äußern Forscher hinsichtlich der Garantie der Fairness die Bedenken, ebenjene führe zu einer Verringerung der Genauigkeit eines Modells [PS22]. Weiter ermöglichen Algorithmen, die mit Daten trainiert wurden, die keine sensiblen Attribute enthalten, höchstwahrscheinlich keine unfaire Behandlung. Lediglich die Auswirkung auf Individuen oder Gruppen führt dennoch möglicherweise zu unbeabsichtigter Diskriminierung [PS22].

Hinsichtlich der Robustheit von Modellen und somit ihrer Resilienz gegenüber Perturbationen, Adversarial Attacks oder Umgebungsveränderungen stehen mehrere Evaluationsansätze zur Verfügung, wie Clever Hans [GP17] und Advbox [Goo+20]. Diese Python Bibliotheken ermöglichen es Forschern, die Robustheit ihrer Modelle möglichst einheitlich zu testen und sind damit für die Forschungsgemeinde von großem Wert. Modelle, die Expertenwissen bei ihren Prädiktionen einbeziehen (informierte ML-Ansätze [Bec+21; Rue+21]), stellen wichtige Schritte für eine Erhöhung der Vertrauenswürdigkeit dar, da sie die Robustheit von Prädiktionen durch Konsistenzprüfungen gegen etabliertes Expertenwissen ermöglichen.

Da gerade bei komplexeren Projekten verteilte Trainingsmethoden verwendet werden, sind die verwendeten Daten und die entstehenden Modelle einem erhöhten Risiko äußerer Einflüsse ausgesetzt [ZJ20]. Dies ist beispielsweise bei Cloud Lösungen der Fall. Insbesondere in solchen Kontexten ist es umso wichtiger, die Integrität der Daten und der Modelle zu garantieren. Diesbezüglich existieren erste Ansätze, die sich mit möglichen Risiken und Lösungen beschäftigen [SEK21; ZJ20]. Eine Analogie zum Integritätsgaranten, den die Nutzung von Blockchain-Technologie darstellt, liegt nahe. Aus diesem Grund entstehen auch in diesem Kontext bereits erste Forschungsarbeiten [Che+18b; Liu+20].

Weiterhin stellt ein verhältnismäßig simpler Schritt der Verlässlichkeitssteigerung die reine Dokumentation und Garantie der Reproduzierbarkeit dar. Dabei handelt es sich etwa um Aufnahmeprotokolle, inklusive Modalitäten und Qualitätsmetriken. Eine vollständige, standardisierte und insbesondere transparente Dokumentation solcher Prozesse erhöht die Reproduzierbarkeit der Ergebnisse anderer Forscher enorm. In Bezug auf DNNs wurden in der einschlägigen Forschung bereits Vorschläge für Taxonomien einer manuellen Dokumentation entwickelt, etwa Fact Sheets [Ric+20] und Model Cards [Mit+19]. Diese Dokumentationstaxonomien beinhalten Informationen hinsichtlich der genutzten Daten, Modelle und Metriken als auch Ziele der

Anwendungen und der Absichten in Bezug auf Verzerrungen und Fairness. Im Europäischen Raum wird darüber hinaus über die Zertifizierung von KI diskutiert, beispielsweise in einem Whitepaper aus dem Jahr 2020 zu ebenjenem Thema [Hee+20]. Die kürzliche Entwicklung des Konzepts der Care-Labels [Mor+21; Mor+22] stellt darüber hinaus eine Schnittstelle zwischen Dokumentationstaxonomie und Zertifizierung dar. Ethisch-rechtliche Auswirkungen der Interpretierbarkeit von DNNs werden ebenfalls untersucht [For+21].

## 2.5 Wissenschaftliche Beobachtung und Quantifikation

Ein viel diskutiertes Zitat, das dem Ökonomeprofessor Peter Ferdinand Drucker zugeschrieben wird, besagt „If you cannot measure it, you cannot manage it.“ [LS16] Doch bereits Galileo Galilei soll gesagt haben „Zähle, was zählbar ist, messe, was messbar ist, und das, was nicht messbar ist, mach messbar.“ [Fin82] Wie in Kapitel 2.1.2 beschrieben, wird die Messung relevanter Zustandsgrößen via Sensorik als Grundlage der Digitalisierung verstanden [Sch+20, S. 18]. Im Rahmen der Nutzung physikalischer Größen und besonders im Kontext von Steuer- und Regelsystemen, ist die Bedeutung ersterer durch physikalische Zusammenhänge gegeben. Für andere Messungen jedoch, wie im Bereich des Managements, ist eine subjektive Einschätzung notwendig, welche die Problematik der Interpretation der Messergebnisse mit sich bringt. In diesem Kontext ist eine Differenzierung einiger Schlüsselbegriffe von Relevanz – Die Begriffe Beobachtung, Quantifikation, Maß, Metrik und Key Performance Indicator:

- Eine wissenschaftliche Beobachtung ist das absichtliche, selektive und aufmerksame Wahrnehmen eines Aspekts eines Prozesses oder Vorgangs [GW97, S. 12]. Die wissenschaftliche Beobachtung bietet die Möglichkeit der geplanten Auswertung und ist geleitet von einer Suchhaltung hinsichtlich eines Aspekts, der für die Auswertung von Interesse ist [GW97, S. 12]. Das Konzept liegt der Wissenschaft als Ganzes zugrunde und wurde in der modernen Forschung besonders im Kontext der Verhaltenspsychologie und Anthropologie formalisiert [CBÖ18, S. 33-34]. Die wissenschaftliche Beobachtung ist der Wegbereiter des Experiments. Ein Beispiel für eine Beobachtung sind die Prädiktionen eines Modells unter Verwendung eines spezifischen Datensatzes.
- Die Quantifikation ist der Prozess, einer Beobachtung einen messbaren Wert zuzuschreiben und sie somit mathematisch abbildbar zu machen [GW97, S. 19]. Die Beobachtung ist dadurch in ihrem Umfang (ihrer Quantität) messbar dargestellt. Ein Beispiel für eine Quantifikation ist etwa das Aufzählen aller detektierter Objekte einer Klasse innerhalb eines Bildes.

- Ein Maß, oder auch eine Messung, ist definierbar als eine quantitative Bestimmung einer physikalischen Größe durch den Vergleich mit einer als Standard angenommenen, festen Größe, wobei das Ergebnis der Messung als numerischer Wert festzuhalten ist [Fer04]. Simpler formuliert ist ein Maß demnach eine Quantität oder ein Grad von etwas, denn wenn eine Beobachtung quantifizierbar ist, so ist sie messbar [Mic97]. Ein Beispiel eines Maßes ist etwa der Prozentsatz der korrekt detektierten Objekte einer Klasse innerhalb eines Bildes.
- Eine Metrik leitet sich aus einer Messung ab und dient dazu, eine Beobachtung zu quantifizieren [Wit18, S. 2]. Eine Metrik trägt somit unter anderem zur Vergleichbarkeit bei. Metriken eignen sich für ihren respektiven Anwendungsfall, wenn sie objektiv, vergleichbar, ökonomisch, valide, zuverlässig, nützlich und normiert sind [Wit18, S. 2-3]. Ein Beispiel einer Metrik ist die in Kapitel 2.2 vorgestellte *IoU*.
- Key Performance Indicators (KPI), auch Leistungsindikatoren, sind Kenngrößen, die der Bewertung quantifizierbarer Metriken dienen, die indikativ für die Performance einer Organisation hinsichtlich eines spezifischen Aspekts sind [Wit18, S. 2]. Ein KPI setzt sich somit aus einer oder mehreren Metriken zusammen. Alle KPI sind Metriken, aber nicht alle Metriken sind KPI. Ein Beispiel für einen KPI ist etwa das EBIT (Earnings Before Interest and Taxes) eines Unternehmens.

Das übergeordnete Ziel der Nutzung all dieser Konzepte ist die Ermöglichung der Operationalisierung. Die Operationalisierung ist der Prozess, der die Metriken festlegt, mithilfe derer ein nicht direkt messbares Phänomen messbar gemacht wird [Emm+16, S. 306-308]. Häufig geschieht dies aus der Inferenz der Messungen anderer, direkt messbarer Phänomene. So ist beispielsweise das Konzept der Qualität ein häufig relevantes, jedoch abstraktes Phänomen, das nicht direkt messbar ist. Aus diesem Grund gilt es andere, messbare Attribute als relevante Attribute auszuwählen und anschließend zu messen. Auf Basis dessen wird letztendlich ein Qualitätsmaß inferiert.

Besonders im Kontext des Qualitätsmanagements sind diverse Ansätze zur Qualitätssicherung entwickelt worden. Manifestiert in der Qualitätsmanagementnorm DIN EN ISO 9000 sind Standards und Anforderungen hinsichtlich der Qualitätssicherung, beispielsweise von betrieblichen Prozessen [DIN 9000:2015-11]. Aus diesem Grund ist eine DIN EN ISO 9000 Zertifizierung in der Industrie weit verbreitet. In diesem Kontext sind Qualitätsmanagementmethoden oder auch Qualitätssicherungstechniken, wie die Failure Mode and Effects Analysis (FMEA) [Sta03], wichtige Werkzeuge zur Bestimmung und Sicherung der angestrebten Qualität. FMEA ist eine systematische Methode zur Bewertung von potenziellen Ausfallmodi und Qualitätsrisiken in einem Prozess oder Produkt. Die Methode analysiert potentielle Ausfälle, ihre Folgen und Ursachen. Darüber hinaus werden Auftretenswahrscheinlichkeit, Bedeutung und Entdeckungswahrscheinlichkeit

der Ausfälle gewichtet dargestellt. Nutzer erhalten durch Anwendung der Methode eine Risk Priority Number (RPN), die als vergleichbare Metrik dient. Basierend auf der RPN werden Risikominderungsmaßnahmen unternommen und nach ihrer Implementierung eine erneute Bestimmung der RPN durchgeführt. Sofern die getroffenen Maßnahmen erfolgreich waren, wird somit eine Minderung der RPN erwartet. Die FMEA Methode wurde seit ihrer Konzeption vielfach weiterentwickelt, kritisiert und adaptiert [LLL13].

Eine alternative Methode, die einen Teil des Quality Function Deployment (QFD) darstellt, ist das House of Quality [HC88]. Das House of Quality ist ein Tool, das ebenfalls im Rahmen der Qualitätssicherung eingesetzt wird und das dazu dient, Kundenanforderungen mit den technischen Umsetzungen im Produkt oder Prozess abzustimmen. Es werden dabei Kundenanforderungen, Designaspekte, Kosten und Umsetzbarkeit gewichtet gegenübergestellt. Basierend auf den Ergebnissen dieser Auswertung werden Maßnahmen definiert, welche die Kundenzufriedenheit steigern und die gewünschte Qualität garantieren.

Als kontinuierliche Verbesserungsmethoden sind die DMAIC (Define, Measure, Analyze, Improve, Control) und PDCA (Plan-Do-Check-Act) Ansätze [SPP10] zu nennen. Beide Ansätze repräsentieren strukturierte Vorgehensweisen, die darauf abzielen, existierende Prozesse kontinuierlich zu analysieren, zu verbessern und zu überwachen, um beispielsweise ihre Qualität oder Effizienz zu steigern. Der DMAIC Ansatz hat seine Wurzeln im Qualitätsmanagement, während der PDCA Ansatz aus dem Lean Management stammt.

Aufgrund des Umstands der nur indirekten Messbarkeit von Konzepten wie dem der Qualität, ist und bleibt die Quantifikation eines solchen Konzepts teils subjektiver Natur. Dennoch gilt es in der Realität oftmals, eine Operationalisierung durchzuführen, die eine möglichst adäquate Approximation bietet [Emm+16, S. 306-308]. Aufgrund der dadurch persistierenden Subjektivität, stellt sich die Frage nach der Evaluation ebenjener Metriken, Definitionen und Konzepte.

Die umfassende wissenschaftliche Evaluation von Metriken, Vorgehensweisen, Definitionen und entwickelten Methoden erfolgt durch diverse Ansätze, die in ihrer Gesamtheit ein tiefgreifendes Verständnis über die Effektivität und Anwendbarkeit der Entwicklungen bieten. Einer der zentralen Bewertungsprozesse ist der Prozess des Peer-Reviews, bei dem Forschungsergebnisse von Experten aus dem relevanten Forschungsgebiet kritisch überprüft werden [Cla03]. Durch diesen Prozess wird nicht nur eine qualitative Einschätzung gewährleistet, sondern auch konstruktives Feedback basierend auf existierender Literatur vermittelt. Dieses Feedback dient der anschließenden Verbesserung und Veränderung der eingereichten Forschungsergebnisse. Der Peer-Review-Prozess ermöglicht die Identifikation von Stärken, Schwächen

und potenziellen Verbesserungsbereichen und dient im Zweifelsfall auch dazu, ungeeignete Forschungsarbeiten zurückzuweisen [SC21].

Ein ebenso bedeutsamer Aspekt der Evaluation ist die Praxisanwendung [GE07]. Hierbei wird Feedback direkt aus der relevanten Industrie oder beruflichen Praxis eingeholt, das auf der tatsächlichen Anwendung der betrachteten Metriken oder Methoden in realen Szenarien basiert. Erfahrungsberichte von Experten aus dem Bereich der Anwendung bieten wertvolle Einblicke in die praktische Relevanz, Umsetzbarkeit und Effizienz. Die Rückmeldungen aus der Industrie tragen dazu bei, die Übereinstimmung der theoretischen Konzepte mit den Anforderungen der realen Welt zu bewerten [Szü18]. Mögliche Limitierungen, die in der Konzeptionalisierungsphase nicht antizipiert wurden, werden letztlich im Kontext der Industrie bemerkt und potentiell gelöst.

Eine fundamentale Rolle in der wissenschaftlichen Evaluationspraxis spielen ebenfalls die empirische Forschung und die Sicherstellung der Reproduzierbarkeit [RS19; Sto+16]. Durch die Durchführung empirischer Studien werden Daten gesammelt und die daraus ermöglichten statistischen Analysen dienen der Bewertung der Zuverlässigkeit und Gültigkeit der betrachteten Metriken oder Methoden. Die Reproduzierbarkeit von Studien ist entscheidend, um die Konsistenz der Ergebnisse unter verschiedenen Umgebungsbedingungen sicherzustellen. In Anbetracht erstellter Definitionen oder Methodiken spielt ebenfalls die Generalisierbarkeit und übergreifende Anwendbarkeit eine Rolle [Car18; Raf+21].

Einen weiteren Evaluationsweg stellt der Vergleich mit etablierten Standards oder Best Practices dar, wobei oftmals eine Überschneidung mit der industriellen Praxis gegeben ist [GF20; Ser+20]. Hierbei wird überprüft, ob die betrachteten Metriken oder Methoden mit anerkannten Industriestandards und -normen übereinstimmen. Dieser Vergleich ermöglicht eine zusätzliche Bewertungsgrundlage hinsichtlich Qualität, Normkonformität und Relevanz. Analog ist in sich formenden Forschungsbereichen auch der Vergleich mit aktuell etablierten state-of-the-art Lösungen ein Evaluationsweg.

In der Summe dieser unterschiedlichen Ansätze wird eine umfassende wissenschaftliche Evaluation empirischer oder qualitativer Forschung gewährleistet. Dieser mehrdimensionale Bewertungsprozess garantiert die Einhaltung von Anforderungen aus Forschung und praktischen Erfordernissen der Industrie seitens der Metriken und Methoden. Dennoch sind und bleiben diese Bewertungen empirischer oder subjektiver Natur.



## Kapitel 3

# Grenzen bisheriger Forschung

**Zusammenfassung** In diesem Kapitel werden die aktuellen Grenzen des Forschungsstands im Kontext der verlässlichen Identifikation dargestellt. Es wird dabei zunächst auf die Anforderungen an Identifikationsverfahren für den Einsatz in der Logistik eingegangen und die damit einhergehende Vision der Identifikation beschrieben. Des Weiteren werden bestehende Identifikationsverfahren und die notwendige Etablierung einer Datengrundlage für die Anwendung in der Logistik besprochen. Zuletzt wird der Mangel an Forschung zur Definition und Quantifikation des Begriffs der Verlässlichkeit dargestellt. Dieses Kapitel motiviert somit die nachfolgende Erarbeitung neuer Ansätze innerhalb der beschriebenen Forschungslücke.

„The dogmas of the quiet past are inadequate to the stormy present.“  
– Abraham Lincoln, Ansprache zur Lage der Union (1862)

### **3.1 Anforderungen an die Identifikation logistischer Entitäten**

Wie im vorangegangenen Kapitel vorgestellt, weisen Identifikationsverfahren im Kontext der Personenüberwachung weiterhin signifikante Fehlerraten und Schwachstellen auf. Dennoch sind sie bereits in der Praxis ihres Anwendungsfeldes anwendbar. Die Anwendbarkeit neuartiger Technologien hängt von den Konsequenzen ihres Versagens im jeweiligen Kontext ab. Für die Anwendung neuartiger technologischer Lösungen im Kontext der Logistik ist eine vorherige Erprobung ebenjener unabdingbar. Oftmals implizieren in der industriellen Anwendung auch nur geringfügige Fehler immense Kosten oder gar Risiken für Leib und Leben. Dem gegenüber steht die Anwendung dieser Methoden für die Identifikation von Individuen im öffentlichen Raum, bei welcher Datenschutzbedenken im Vordergrund stehen.

In diesem Sinne ist die bislang mangelnde Genauigkeit und Verlässlichkeit der existierenden Identifikationsverfahren als Ausschlusskriterium für die Anwendung in der logistischen Praxis zu verstehen. So ist für eine effiziente Anwendung ein fundierteres Verständnis (im Sinne der Erklärbarkeit) über die Funktionsweise des Identifikationsmodells nötig. Dieses würde es einem Nutzer ermöglichen, an entscheidenden Stellen im Entscheidungsprozess manuell einzugreifen. Alternativ gilt es, die Funktionsweise des Modells deutlich einzuschränken und auf einen spezifischen Anwendungsfall und Datensatz zu fokussieren, um zufriedenstellende Identifikationsgenauigkeiten zu ermöglichen. Darüber hinaus sind die Echtzeitfähigkeit und die Kosteneffizienz eines solchen Systems in vielen Fällen von äußerster Relevanz. Die bisher angebotenen Lösungen, die extrinsische Marker verwenden, sind aufgrund des in der Industrie unausweichlichen Verschleißes, nicht langfristig nutzbar und implizieren Einschränkungen in ihrer Einsetzbarkeit.

Basierend auf den in diesem Abschnitt beschriebenen Anforderungen hat eine industrielle Identifikationslösung demnach gewisse Charakteristika aufzuweisen, um sinnvoll angewendet zu werden und auf Akzeptanz zu stoßen. Der aktuelle Forschungsstand bietet eine solche Lösung nicht, weswegen eine eigenständige Entwicklung einer solchen Identifikationslösung notwendig ist.

### **3.2 Bestehende Datengrundlage für die Identifikation logistischer Entitäten**

Datensätze aus Industrie und Forschung existieren bereits zur Genüge. Diese Datensätze werden jedoch für die Klassifikation oder Detektion von logistischen Entitäten verwendet. Somit ist die mehrfache Aufzeichnung und Annotation der jeweiligen Entitäten in den Datensätzen nicht garantiert, wo-

durch die Nutzung zum Zweck der Identifikation meistens nicht ermöglicht wird.

Als Beispiel eines solchen Datensatzes ist der Datensatz LOCO (Logistics Objects in COntext) der Technischen Universität München zu nennen [May+20]. Dieser Datensatz enthält Bilder von Paletten, Kleinladungsträgern, Gabelstaplern, Gitterboxen und Handgabelhubwägen, die teilweise annotiert sind. Die Bilder wurden in der Bewegung aufgenommen, mit einer auf einem Wagen montierten Kamera. Es ist bei der Aufnahme des Datensatzes nicht bekannt, ob eine der Entitäten mehrmals aufgenommen wurde und somit wurde jede ID nur einmalig vergeben, auch wenn die respektive Entität womöglich mehrmals aufgenommen worden ist.

Einen ebenfalls industriebezogenen, jedoch synthetischen Datensatz stellt SORDI (Synthetic Object Recognition Dataset for Industries) dar [Aka+22]. Der im Jahr 2022 veröffentlichte Datensatz entstand aus einer Kollaboration von BMW, Nvidia und Microsoft und stellt synthetische Daten zur Detektion logistischer Entitäten dar. Der mit generativen Modellen und Nvidia Omniverse erstellte Datensatz beinhaltet Instanzen von Gabelstaplern, Paletten und Gitterboxen. Da die Daten mittels 3D-Rendering generiert worden sind, ermöglichen sie über die Detektion auch eine Segmentierung der Entitäten. Die enthaltenen Entitäten werden aus vordefinierten Modellen in synthetischen Umgebungen eingefügt und sind somit intragruppal voneinander ununterscheidbar, sofern nicht einzigartige Texturen ergänzt werden.

Einen weiteren industriellen Datensatz stellt ITD (Industrial Tools Detection) dar [Luo+19]. Dieser Datensatz beschäftigt sich mit der blickpunktevarianten Darstellung industrieller Werkzeuge, insbesondere für den Zweck der automatisierten Greifrobotik in der industriellen Anwendung. Es wurden dabei 24 Werkzeugklassen ausgewählt, wie Scheren, Maßbänder, Zangen und Feilen, die für die Aufgabe der Detektion verwendet werden. Eine Unterscheidung der jeweiligen Werkzeuge innerhalb der Werkzeugklassen ist jedoch mangels Annotation nicht möglich, analog zu den bereits genannten industriellen Datensätzen.

Wie eingangs beschrieben und anhand der genannten Datensätze illustriert, sind industrielle Datensätze nicht zwangsläufig zum Zweck der Identifikation spezifischer Entitäten nutzbar. Im Kontrast dazu sind selbstverständlich auch Datensätze spezifisch im Kontext der Forschung der Identifikation erstellt worden. Bereits in den 2000er Jahren wurde mit ETH3 [ELV07] (namensgebend war die ETH Zürich) ein Datensatz zur Detektion von Personen vorgestellt, der als erster Schritt in Richtung der Identifikation zu betrachten ist. Zur selben Zeit wurde auch VIPeR [GBT07] (Viewpoint Invariant Pedestrian Recognition) veröffentlicht, der als erster Benchmarkdatensatz für die Forschungsgemeinde gilt. Der Datensatz umfasst 632 Personen, aufgenommen mit zwei Kameras unter verschiedenen Blickwinkeln und Lichtbedingungen, wobei jede Person pro Kamera ein Mal aufgenommen wurde.

In den 2010er Jahren nahm die Komplexität der Datensätze signifikant zu und so bieten Datensätze wie CUHK03 [Li+14] (Chinese University of Hongkong) und Market-1501 [Zhe+15] deutlich mehr Bilder, als es ihre Vorgänger taten. CUHK03 beinhaltet 1.360 Personen, aufgezeichnet mit sechs Kameras auf einem Universitätscampus. Jede Person wurde dabei von zwei der Kameras aufgenommen. Der wohl etablierteste Benchmarkdatensatz, Market-1501, beinhaltet eine ähnlich hohe Anzahl an Personen, nämlich die namensgebende Anzahl von 1.501 Personen, die auf dem Campus der Tsinghua Universität aufgenommen wurden. Der Datensatz basiert jedoch auf der deutlich größeren Menge von 32.668 Bildern, da jede Person mehrmals von jeder der sechs verwendeten Kameras aufgenommen worden ist. Darüber hinaus werden über eine halbe Million Distraktorbilder bereitgestellt. Die Größe und die hohe Anzahl an Bildern pro Personen unterscheidet Market-1501 entschieden von seinen Vorgängern und hat den Datensatz somit als state-of-the-art Datensatz etabliert.

**Tabelle 3.1** Identifikations- und Industriedatensätze und ihre Modalitäten.

Kategorie	Datensatz	Jahr	IDs	Kameras	Bilder	Auflösung [px]
Detektion & Identifikation	ETH3 [ELV07]	2007	-	1	2.293	Variiert
	VIPeR [GBT07]	2007	632	2	1.264	128 × 48
	CUHK03 [Li+14]	2014	1.360	6	13.164	Variiert
	Market-1501 [Zhe+15]	2015	1.501	6	32.668	128 × 64
	DukeMTMC [Ris+16]	2016	2.834	8	36.441	Variiert
Industrie	ITD [Luo+19]	2019	-	1	11.000	1024 × 575
	LOCO [May+20]	2020	-	2	37.988	Variiert
	SORDI [Aka+22]	2022	-	-	200.000	1280 × 720

Zuletzt ist der Datensatz DukeMTMC [Ris+16] (Duke University Multi-Camera Multi-Tracking) zu nennen, der 2.834 Personen, aufgenommen von acht Kameras enthält. Der Datensatz beinhaltet in Summe 36.441 Bilder und ist somit rein numerisch noch umfangreicher als Market-1501 [Ris+16]. DukeMTMC ist allerdings auch Teil einer Datenschutzkontroverse, aufgrund derer der Datensatz heutzutage nicht mehr verwendet wird und von den Autoren und der namensgebenden Duke University zurückgezogen wurde. Dies ist der Fall, da die auf dem Campus der Duke University erstellten Aufnahmen ohne die Kenntnis der darin enthaltenen Studenten aufgezeichnet und verbreitet wurden [PMN21]. Damit ist DukeMTMC letztlich ein exzellentes Beispiel für einen Datensatz, der einen problematischen Umgang mit sensiblen Daten in Form von Persönlichkeitsrechten darstellt. Somit spielt

DukeMTMC auch im Hinblick der sicheren und fairen Nutzung von Daten im Kontext des Maschinellen Lernens eine Rolle [Rut+24b].

Auch über den Menschen als Subjekt hinaus gibt es Datensätze, die für den Zweck der Identifikation erstellt worden sind. Ein Beispiel dafür sind Datensätze für die Identifikation von Tieren [Čer+24]. Die zuletzt genannte Gruppe der Datensätze erfüllt insofern die Ansprüche an die Aufgabe der Identifikation. Jedoch sind diese Datensätze beschränkt auf Subjekte, die keine logistischen Entitäten sind. Diese Einschränkung stellt einen Teil der Forschungslücke dar, welche diese Arbeit schließt. Eine Übersicht der relevanten, besprochenen Datensätze ist Tab. 3.1 zu entnehmen.

Erschwerend zu den Einschränkungen der vorgestellten Datensätze kommen im Kontext der industriellen Nutzung häufig Sorgen in Bezug auf Datenschutz und konkurrierende Unternehmen hinzu. Somit existieren womöglich Datensätze, welche die beschriebenen Anforderungen erfüllen, jedoch nicht öffentlich zugänglich sind. Unter anderem aus diesem Grund gewinnt die Erzeugung synthetischer Daten an Relevanz [Mat+23; Rut+22b] und ist in einigen Fällen die einzige Möglichkeit, um geeignete Modelle zu trainieren, was beispielsweise die Entwicklung des Datensatzes SORDI bestärkt. Das Sinken der Prädiktionsgenauigkeit der Modelle unter Nutzung größerer Datensätze ist zu erwarten [ZYH16], weshalb die Erstellung diverser und möglichst großer Datensätze von Interesse ist. Zusammenfassend lässt sich demnach die zwingende Notwendigkeit der Erstellung eigener Datensätze zur Identifikation logistischer Entitäten konstatieren. Diese Datensätze gilt es dabei möglichst für eine industrielle Nutzung geeignet zu gestalten und systematisch und strukturiert zu erstellen und zu dokumentieren.

### 3.3 Bestehende Identifikationsverfahren

Wie in Kapitel 2.2.2 dargestellt, existieren diverse Identifikationslösungen für unterschiedlichste Subjekte und Entitäten. Die dafür verwendeten Modelle sind jedoch aufgrund ihrer Trainingsdaten an ihren spezifischen Anwendungsfall gebunden. Dennoch sind bestehende Lösungen in Form von vortrainierten Modellen weiterverwendbar. So sind im Kontext der Identifikation von Personen einige Modelle entwickelt worden, die eine hohe Identifikationsgenauigkeit auf den im vorherigen Kapitel genannten Benchmarkdatensätzen vorweisen. Diese Modelle basieren in der Regel auf Detektions- oder Klassifikationsmodellen, die bereits in ihren jeweiligen Disziplinen etabliert sind.

So ist etwa MobileNetV2 [San+18] ursprünglich ein Klassifikationsmodell, das für seine niedrigen Hardwareanforderungen bekannt ist, auf gängige Identifikationsdatensätze angewendet worden und erzielte trotz seiner niedrigen Komplexität vielversprechende Ergebnisse [AD21]. Seine Namensgebung ist angelehnt an die potentielle mobile Nutzung des Modells, die durch

die niedrigen Hardwareanforderungen des Modells ermöglicht wird. Diese entstehen etwa durch die invertierte Nutzung von Residual Layers (Layers, die durch Hinzufügen einer Skip Connection den direkten Fluss von Informationen durch das Modell ermöglichen), gefolgt von einem linearen Bottleneck (im Gegensatz zu einer gängigen Aktivierungsfunktion wie ReLU).

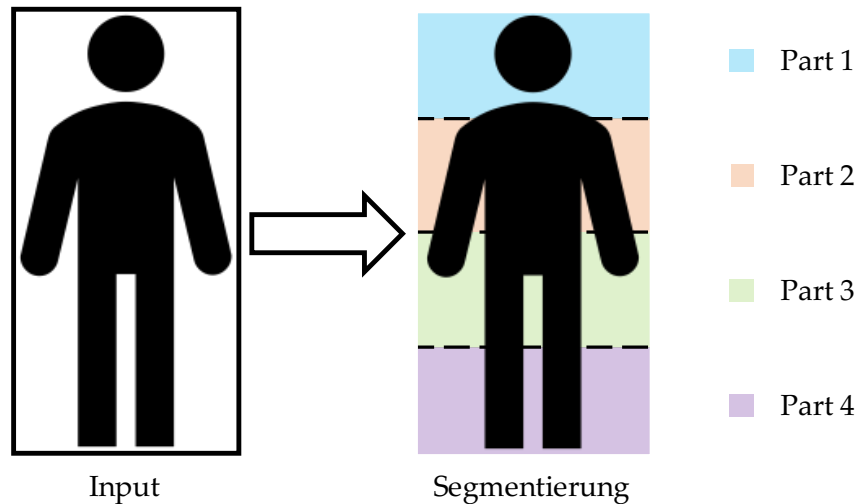


Abb. 3.1 Visuelle Veranschaulichung des Part-Based Convolutional Baseline Ansatzes.

Ein weiteres, weitläufig genutztes Modell stellt PCB (Part-Based Convolutional Baseline) dar. PCB teilt seine Inputbilder in vier bis sechs Ebenen (Parts) auf, um Personen in Bildern zu segmentieren (etwa in Kopf, Torso, Beine, Füße), wie in Abb. 3.1 dargestellt ist. Das auf ResNet50 basierende Modell verwendet mehrere Teilnetze, um die verschiedenen Teile des Bildes zu verarbeiten und erstellt dann eine aggregierte Darstellung für die Identifikation [Sun+18]. PCB unterscheidet sich von ResNet50 in seinen letzten Layern, in welchen das Global Average Pooling durch ein Average Pooling Layer ersetzt wird und das Modell durch ein Dense Layer und eine Softmax Aktivierungsfunktion komplettiert wird.

Im Gegensatz zu MobileNetV2 und PCB wurde das Modell OSNet (Omni-Scale Network) [Zho+19b] spezifisch für die Identifikation von Personen entwickelt. Auch OSNet basiert auf der ResNet Architektur und ist wie MobileNetV2 als light-weight Modell entwickelt worden. OSNet erzielt im Gegensatz zu MobileNetV2 jedoch kompetitivere Ergebnisse.

Die meisten Identifikationsansätze berücksichtigen räumliche oder zeitliche Beschränkung nicht. So wird ein Query Bild mit allen Bildern der Galerie abgeglichen, ihre jeweilige Distanz im Embedding Space bestimmt, ein Ranking der Distanzen erstellt und das Bild mit der kleinsten Distanz als Prä-

diktion ausgegeben. Das Modell st-ReID (Spatial Temporal Re-identification) [Wan+19], ist das erste etablierte Modell, das auch räumlich-zeitliche Faktoren in Betracht zieht. So wird etwa bei einer Nutzung mehrerer Kameras, die räumliche Distanz der Kameras zueinander bei der Wahrscheinlichkeit der Zuordnung einer ID mit einbezogen. Der Vorteil dieses Vorgehens ist das a priori Filtern des Datenraums, was eine effizientere und akkuratere Identifikation ermöglicht.

Neben der adaptierten ResNet Architektur, die für PCB und OSNet verwendet wird, wurde auch ResNet101 erfolgreich für die Identifikation von Personen eingesetzt [Fu+21]. Die Verwendung von Residual Layers ist über verschiedene Modelle wie MobileNetV2, PCB und ResNet101, neben der Klassifikation somit auch im Bereich der Identifikation von Personen etabliert. ResNet101 unterscheidet sich von ResNet50 hinsichtlich seiner namensgebenden, architektonischen Tiefe – Das Modell verfügt über 101 Layers.

**Tabelle 3.2** Gängige Identifikationsmodelle und ihre Ergebnisse bei Anwendung auf den Datensatz Market-1501.

Modell	Jahr	Rank-1-Accuracy [%]	mean Average Precision [%]
MobileNetV2 [San+18]	2018	49,8	25
PCB [Sun+18]	2018	92,3	77,4
OSNet [Zho+19b]	2019	62,8	35
st-ReID [Wan+19]	2019	98,1	87,6
ResNet101 [Fu+21]	2021	95,8	89
LA-Transformer [SKC21]	2021	98,27	94,46

Die Entwickler des LA-Transformer (Locally Aware Transformer) argumentieren, Vision Transformer Modelle würden in der nahen Zukunft gängige Convolutional Neural Networks ersetzen [SKC21]. Transformer Modelle unterscheiden sich gegenüber konventionellen Modellen des Maschinellen Lernens, durch die Nutzung von Attention Layers statt Convolutional Layers. Diese Layers ermöglichen eine Gewichtung spezifischer Teile des zu verarbeitenden Bildes. Als Teil dieser Entwicklung hin zu einer mehrheitlichen Nutzung von Transformer Modellen sehen die Autoren das LA-Transformer Modell. Entgegen der gängigen Praxis von Vision Transfer Modellen werden hierbei lokale statt globale Klassifikations-Tokens ausgegeben. Im Kontrast zu globalen Tokens und analog zu den Pools der Convolutional Neural Networks enthalten diese Tokens zusätzliche, detaillierte Informationen über lokale Bereiche des Bildes. Dieses Modell und die Unterteilung des Gesamtbildes lehnt sich laut der Autoren an den Ansatz des PCB Modells an. Eine

Übersicht der genannten Modelle und ihrer Performance auf dem Datensatz Market-1501 ist Tab. 3.2 zu entnehmen.

Bis auf MobileNetV2 und OSNet sind die vorgestellten Modelle rechenintensiv und somit nicht mit eingeschränkter Hardware nutzbar. Eine Modellart, die ebenjenes ermöglicht, sind Graph Neural Networks (GNN). Bei der Nutzung von GNNs wird eine Repräsentation basierend auf Graphen, analog zur gängigen Vektorrepräsentation geschaffen. Diese sind in der Regel niedrigdimensionaler als ein Featurevektor, weshalb sie weniger Hardwareressourcen benötigen. Die bisherigen Lösungen im Bereich der Identifikation, die Graphen verwenden, sind jedoch spezifisch für die Nutzung im Kontext der Personenidentifikation entwickelt worden. Sie dienen etwa dazu, die Gelenke der Personen mit Knoten und die dazwischenliegenden Gliedmaßen mit Kanten darzustellen [KMS22].

Außerhalb der Anwendung der Identifikation von Personen sind auch Modelle zu finden, die spezifisch auf Bildern logistischer Entitäten trainiert worden sind. So haben Li et al. eine Adaption des VGGNet (Visual Geometry Group) Models zur Detektion von Paletten trainiert [Li+19]. Unter Verwendung von Tiefeninformationen entwickelten darüber hinaus Molter et al. [MF18] und Varga et al. [VN16] Lösungen für die Lokalisation und Detektion von Paletten. Keine dieser Lösungen ermöglicht jedoch die Identifikation der Paletten, wie sie im Rahmen dieser Arbeit gefordert ist.

Zusammengefasst mangelt es angesichts der präsentierten Modelle an generalisierbaren Lösungen, die unter eingeschränkter Hardware nutzbar sind. Darüber hinaus sind keine Modelle identifiziert worden, die spezifisch für die Identifikation logistischer Entitäten trainiert und entwickelt worden sind.

### 3.4 Verlässlichkeitsbegriff im Maschinellen Lernen

Im Vergleich zu anderen Forschungsbereichen ist das Maschinelle Lernen noch eine junge Domäne. Damit einher geht die Nutzung von Schlüsselbegriffen, deren vereinheitlichte Definition teilweise noch nicht abgeschlossen ist. Das betrifft ebenfalls standardisierte Vorgehen und Praktiken. Darüber hinaus fehlt jedoch in einigen Subdomänen gar der Konsens bezüglich der Definition relevanter Begrifflichkeiten. So wird im Kontext der erklärten Künstlichen Intelligenz (XAI) der Begriff der Verlässlichkeit (Trust) weitgehend uneinheitlich genutzt und ist nicht definiert. Die Nutzung dieses Begriffs geht mit weiteren, mehrdeutig verwendeten Begriffen einher, auf welche im Folgenden eingegangen wird. Der Stand der Forschung zu diesen Begrifflichkeiten wird deshalb als noch unabgeschlossener wissenschaftlicher Diskurs verstanden.

Es existieren bereits erste Versuche, generalisierte Richtlinien für die erklärbare Nutzung von Systemen des Maschinellen Lernens zu gewährleisten

[SL18]. Es wird ebenfalls die Adaptierung und Anwendung von Modellen aus der Soziologie, wie das ABI-Modell (Ability, Benevolence, Integrity) [MDS95] untersucht. Damit einhergehend werden auch in den Rechtswissenschaften Vorschriften, Gesetzesentwürfe und Auditingverfahren untersucht und erarbeitet [Lep+18; Raj+20; YHP20]. Im Rahmen dieser Richtlinien ist weiterhin der Dokumentationsprozess von Relevanz und ein wichtiger erster Schritt [Raj+20] hin zu einem verlässlichen Vorgehen.

Während diese Prozesse versuchen, eine (rechtliche) Struktur in den noch jungen Forschungsbereich zu bringen, mangelt es allerdings weiterhin an klaren, standardisiert genutzten Definitionen grundlegender Begrifflichkeiten. So werden vielfach Begriffe synonym genutzt, wie die Begriffe der Erklärbarkeit (Explainability) und Verständlichkeit (Comprehensibility) [Her+21], der Abhängigkeit (Reliance) und der Verlässlichkeit (Trust) [Rya20], oder der Interpretierbarkeit (Interpretability) und wieder einmal Erklärbarkeit (Explainability) [Her+23]. Um eine eindeutige Kommunikation in der Forschungsgemeinde zu ermöglichen, ist deshalb eine vereinheitlichte Nutzung und Taxonomie dieser Begriffe nötig und ihrer Zusammenhänge notwendig. Dabei liegt im Rahmen dieser Arbeit der Fokus auf dem Begriff der Verlässlichkeit. Der Zusammenhang der Verlässlichkeit zu Begriffen wie den oben genannten ist dabei festzustellen.

Die in Kapitel 2.4 dargelegten Themenschwerpunkte der Forschung im Bereich der Verlässlichkeit des Maschinellen Lernens und ersten Schritte hin zu einer Definition ebenjener weisen klar die Grenzen des bisherigen Forschungsstands auf – eine einheitliche Definition des Begriffs, oder gar zusammenhängender Begriffe wie den oben genannten, existiert aktuell noch nicht. Basierend auf diesen Limitierungen ist die Ambition dieses Werks, die Erarbeitung einer operationellen Definition der Verlässlichkeit, basierend auf dem herrschenden Verständnis der Verlässlichkeit in der aktuellen Forschung.

Auch wenn somit die Erarbeitung einer solchen Definition qualitativer Natur ist, gilt es dennoch Resultate mit greifbarem Nutzwert zu garantieren. Roskam et al. fassen dieses Bestreben mit den Worten „[...] definitions must be casted in terms of observable behavior“ [Ros89] zusammen. Konkret bedeutet das im Kontext der industriellen und wissenschaftlichen Nutzung einer solchen Definition eine Operationalisierung, wie sie in Kapitel 2.5 aufgezeigt wurde. Diese basiert demnach auf der Erarbeitung definierbarer und messbarer Metriken, auf Basis derer sich letztlich die Verlässlichkeit bestimmen und quantifizieren lässt.

Auch im Hinblick auf eine Verlässlichkeitsmetrik jedoch, hat sich noch keine Standardisierung oder mehrheitliche Nutzung in der Forschungsgemeinde etabliert. Erste Vorschläge zur Messung einzelner Phänomene, die als Aspekte der Verlässlichkeit zu verstehen sind, wurden bereits erarbeitet (beispielsweise die False Discovery Rate (FDR) [Sto11]). Diese erfüllen

allerdings nicht den gewünschten Zweck der holistischen Evaluation eines Modells oder Datensatzes hinsichtlich seiner Verlässlichkeit. Eine Adaptierung von bereits entwickelten und anwendbaren Metriken, beispielsweise aus dem Bereich der Qualitätssicherung (siehe Kapitel 2.5, etwa QFD, FMEA), wurde bislang ebenfalls nicht etabliert.

Aus diesem Grund ist die Erarbeitung einer neuen, generalisierten Metrik der Verlässlichkeit nötig, die sich auf Best Practices stützt und einen ersten Vorschlag für die Forschungsgemeinde darstellt. Bei der Erstellung einer solchen Metrik gilt es zu bedenken, wie eine sinnvolle Evaluierung auszugestalten ist, um einen logischen Kreisschluss zu vermeiden. Es gilt also sinngemäß, angelehnt an Edwin Borings umstrittene [VKB14] Beschreibung des Intelligenzquotiententests [Bor61], das Resultat ‚Verlässlichkeit ist, was die Verlässlichkeitsmetrik misst‘ zu vermeiden. Auch wenn es legitim ist, das Ergebnis einer solchen Metrik oder eines solchen Tests lediglich als eben das zu deklarieren – nicht mehr und nicht weniger als das Ergebnis des Tests – kann die Güte der Evaluation als solche weiterhin untersucht werden. Der Kern der Operationalisierung ist es nämlich, eine präliminäre Lösung zu ermitteln, mithilfe derer eine erste approximative Messung eines nicht direkt messbaren Phänomens ermöglicht wird. Konzeptionelle Operationalisierungen formalisieren insofern möglichst präzise eine subjektive Beobachtung [Ros89]. Werke wie jene von Dmitriev und Wu [DW16] bieten Ansätze zur Evaluation von Metriken und gängige Untersuchungen stellen dabei die Reliabilität, also die Konsistenz der Ergebnisse der Anwendung je Nutzer, in den Vordergrund. Zusammenfassend lässt sich die Notwendigkeit der Erarbeitung einer Definition und Metrik der Verlässlichkeit im Kontext des Maschinellen Lernens festhalten, da sowohl hinsichtlich der Definition, als auch der Quantifikation des Begriffs wenige Vorarbeiten existieren.

### 3.5 Implikationen der Forschungslücke für das entwickelte Vorgehen

Es leiten sich aus der beschriebenen Forschungslücke die folgenden Anforderungen für die wissenschaftliche Bearbeitung der Forschungsfragen ab:

- Die Notwendigkeit der Auswahl logistischer Entitäten und der Datensatzerstellung. Die in der bisherigen Forschung erstellten Datensätze sind entweder nicht für die Re-identifikation nutzbar oder beinhalten keine logistischen Entitäten. Aus diesem Grund ist es notwendig, eine Datengrundlage für Identifikationsprobleme zu erstellen. Zur Auswahl stehen dafür standardisierte Entitäten, deren Oberflächenmerkmale sich für ein visuelles Identifikationsverfahren eignen.

Ausgewählt werden in dieser Arbeit deshalb Euro-Paletten sowie Stapeltransportkisten. Diese weitläufig genutzten logistischen Entitäten eignen sich aufgrund ihrer einzigartigen Oberflächenbeschaffenheiten. Im Fall der Euro-Paletten handelt es sich dabei um den Pressspan, aus welchem die Klötze der Paletten bestehen. Im Fall der Stapeltransportkisten handelt es sich um den galvanisierten Stahl, aus welchem die Kisten hergestellt sind. Durch die Galvanisierung wird der Stahl zwecks Korrosionsschutz mit einer Zinkschicht versehen. Bei der Erstarrung des flüssigen Zinks entstehen dabei auf der Oberfläche des Materials kristalline Strukturen (Zinkblumen), die jede verzinkte Oberfläche visuell einzigartig machen.

- Die Notwendigkeit der Adaptierung bestehender und der Erstellung neuer, geeigneter Identifikationsmodelle. Unter der Nutzung geeigneter Trainingsdaten liegt die Verwendbarkeit bestehender Architekturen für die Identifikation logistischer Entitäten nahe. Dennoch ist die Echtzeitnutzung komplexer Modelle wie PCB von der Verfügbarkeit leistungsstarker Hardware abhängig. Im Falle der Identifikation von Personen im öffentlichen Raum werden im Regelfall statische Kamerasysteme verwendet, die ihr Bildmaterial an einen Server weiterleiten, auf welchem die letztendliche Rechenarbeit stattfindet. In den seltensten Fällen wird dabei eine Identifikation in Echtzeit benötigt. Bei der Nutzung dieser Methoden in einem industriellen Kontext hingegen ist die Notwendigkeit der Echtzeitfähigkeit häufig gegeben und die verfügbare Hardware dennoch eingeschränkt. Die benötigte Datenmengen für industrielle Anwendungen fällt außerdem deutlich größer aus als bei der Identifikation von Personen (siehe Datensatzgrößen in Kapitel 3.3). So ist auch die Implementierung einer effizienten Datenbank von Vorteil, etwa hinsichtlich des Formats, in welchem die Entitäten hinterlegt werden oder hinsichtlich des Verfahrens, mit welchem die Datenbank durchsucht wird.

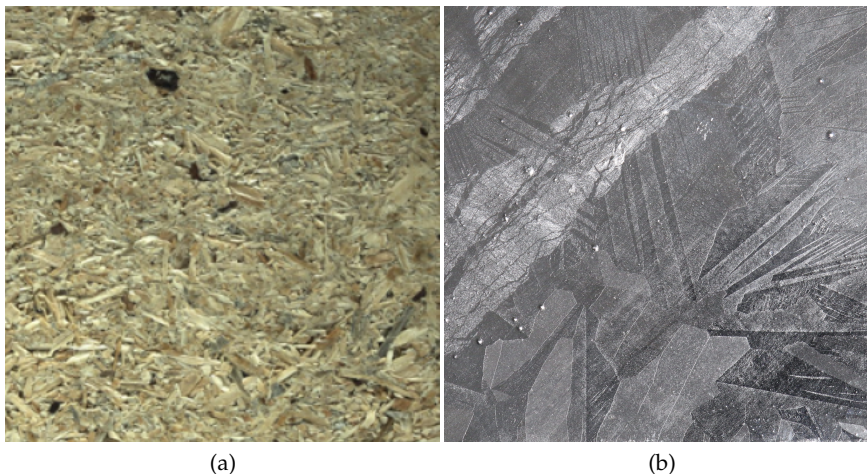
Aus diesem Grund werden sowohl eine state-of-the-art Architektur aus der Identifikation von Personen (PCB) auf neuen, industriebezogenen Daten trainiert und getestet, als auch neue Algorithmen zur Identifikation logistischer Entitäten vorgeschlagen. Letztere verwenden Graphrepräsentationen, im Gegensatz zu den Vektorrepräsentationen, die klassischerweise im Deep Learning verwendet werden. Dies ermöglicht eine niedrigere Dimensionalität der gespeicherten Daten, die sich besonders für die Nutzung unter eingeschränkter Hardware eignet. Dieser Ansatz soll weiterhin in Kombination mit rudimentären Modellen als Ensemble verwendet werden.

- Die Notwendigkeit einer eindeutigen Definition und Quantifikationsmetrik für den Begriff der Verlässlichkeit im Kontext des Maschinellen Lernens. In der Forschung herrscht bislang Uneinigkeit hinsichtlich relevanter Begrifflichkeiten wie Fairness, Erklärbarkeit, Robustheit, Datenschutz

und nicht zuletzt Verlässlichkeit. Diese terminologische Ambiguität erhöht den Schwierigkeitsgrad des Austauschs zwischen Forschern. Aus diesem Grund wird in dieser Arbeit eine operationelle Definition vorgeschlagen, welche die Schlüsselbegriffe im Kontext der Verlässlichkeit des Maschinellen Lernens vereint. Der Mangel einer solchen Definition hindert außerdem die Forschungsgemeinschaft daran, eine Metrik zu etablieren, mithilfe derer die Verlässlichkeit im Kontext des Maschinellen Lernens quantifiziert wird. Aus diesem Grund wird weiterhin, aufbauend auf der erarbeiteten Definition, eine Metrik der Verlässlichkeit erstellt, die dem Nutzer ein numerisch greifbares und vergleichbares Verständnis für die Verlässlichkeit eines Datensatzes oder eines Modells bietet.

### 3.6 Abgrenzung des Forschungsgegenstands

Identifikationsverfahren sind für verschiedenste Zwecke einsetzbar. Gerade deshalb gilt es abzugrenzen, welche Aspekte im Rahmen dieser Arbeit nicht betrachtet werden. Wie in Kapitel 2 beschrieben werden innerhalb dieser Arbeit Entitäten, spezifischer logistische Entitäten betrachtet. Dieser Begriff grenzt sich klar von dem Begriff der Logistikobjekte nach [Gud04] ab, welcher auch Lebewesen als mögliches Logistikobjekt beinhaltet. Somit sind im Rahmen dieser Arbeit Menschen oder Lebewesen im Allgemeinen nicht Teil des Betrachtungsrahmens.



**Abb. 3.2** Beispiele einzigartiger, visueller Oberflächenmerkmale eines äußeren Klotzes einer Euro-Palette (a) und der Oberfläche einer Stapeltransportkiste (b), wie sie auch in Abb. 2.1 zu sehen sind.

Des Weiteren beschäftigt sich diese Arbeit mit der Identifikation basierend auf inhärenten visuellen Merkmalen ebenjener logistischer Entitäten, wie sie in Abb. 3.2 zu sehen sind. Es findet demnach keine Identifikation basierend

auf chemischen Eigenschaften, Oberflächentopologien, Kontextinformationen, extrinsischen Markern oder anderen Charakteristika der Entitäten statt.

Das deklarierte Ziel ist es ebenfalls, das erarbeitete Vorgehen zur Identifikation leicht zugänglich und in der Praxis anwendbar zu machen. Aus diesem Grund wird sich im Rahmen dieser Arbeit auf die Nutzung simpler technischer Methoden beschränkt. RGB-Kameras mit gängigen Auflösungen sind somit für den Anwendungsfall ausreichend, da Oberflächenmerkmale verwendet werden, die mit dem bloßen Auge erkennbar sind. Die verwendeten Kameras nehmen somit lediglich die Oberfläche der betrachteten Entitäten auf, erzeugen aber nicht mikroskopische Nahaufnahmen, wie in es in anderen Arbeiten der Fall ist [TKI17; Wig20].

Da der Fokus des Vorgehens auf der industriellen Anwendung liegt, werden Daten und Versuchsaufbauten verwendet, die ebenjenem industriellen Kontext gleichen. Somit werden Randfälle, etwa hinsichtlich der Beleuchtung oder der Betrachtungswinkel, nicht betrachtet. Es wird im Kontext der Robustheit lediglich von semi-kontrollierten industriellen Umgebungen ausgegangen.

Die primäre Erkenntnisquelle dieser Arbeit, hinsichtlich der Genauigkeit der Identifikationsverfahren und ihrer Verlässlichkeit, ist der Empirismus. Die sekundäre Erkenntnisquelle ist die qualitative Evaluation der Verlässlichkeitsdefinition und ihrer Reliabilität. So werden Versuche ausgeführt, auf deren Basis sich die schlussendlichen Erkenntnisse der Arbeit stützen.

Da es eine der Ambitionen dieser Arbeit ist, einen Teil zur Forschung hinsichtlich der Vision des Lagers der Zukunft beizutragen (siehe Kapitel 2.1.4), werden aufgrund der interdisziplinären Natur dieses Themas diverse Disziplinen bearbeitet. Auch der in Vorbereitung stehende Sonderforschungsbereich „Datengetriebene Logistik für die Industrie 4.0“ der Technischen Universität Dortmund spielt hierbei als Motivator eine einordnende Rolle. Der Erkenntnisgewinn dieser Arbeit dient dem Fortschritt hinsichtlich der Realisierung der Vision des Lagers der Zukunft, wodurch die experimentellen Einschränkungen der Betrachtungen dieser Arbeit begründet sind.



## Kapitel 4

# Vorgehen zur verlässlichen Identifikation logistischer Entitäten

**Zusammenfassung** In diesem Kapitel wird das zentrale Vorgehen zur verlässlichen Identifikation logistischer Entitäten vorgestellt. Dabei wird das Vorhaben zunächst in der bestehenden Forschung eingeordnet und eine wissenschaftstheoretische Positionierung formuliert. Weiter wird der Identifikationsprozess bezüglich des Modelltrainings, der Datensatzerstellung und der Evaluation der Ergebnisse beschrieben. Zuletzt werden die relevanten Aspekte der Verlässlichkeit im Maschinellen Lernen hergeleitet und das Evaluationsvorgehen im Kontext der Verlässlichkeitsbestimmung präsentiert.

„[...] la chance ne sourit qu’aux esprits bien préparés.“  
– Louis Pasteur, Eröffnung der naturwissenschaftlichen Fakultät der Universität Lille (1854)

## 4.1 Erstellung geeigneter Datensätze für die Identifikation logistischer Entitäten

Wie eingangs beschrieben, mangelt es im jungen Forschungsfeld des Maschinellen Lernens an einem standardisierten Datensatzerstellungsvorgehen. Dennoch gibt es Best Practices und weitreichend genutzte und etablierte Datensätze, an denen Forscher sich bei der Erstellung eines eigenen Datensatzes orientieren. Der allgemeine Anspruch an Datenqualität, wie er etwa von Pipino et al. beschrieben wird [PLW02], ist ebenfalls auf Datensätze für das Maschinelle Lernen anwendbar. Demnach sind Charakteristika wie Zugänglichkeit, Umfang, Objektivität, adäquate Repräsentation, Vollständigkeit, Fehlerfreiheit und Interpretierbarkeit auch im Kontext dieser Arbeit von Relevanz. Dennoch obliegt die Beurteilung ebenjener Charakteristika häufig den Nutzern eines Datensatzes. Folglich ist die Beurteilung einiger Charakteristika subjektiver Natur (etwa bezüglich der adäquaten Repräsentation), während jedoch andere Charakteristika objektiv quantifizierbar (etwa der Umfang) aber nicht objektiv bewertbar sind (etwa ob der Umfang als ‚groß‘, ‚klein‘, ‚zu wenig‘ oder ‚ausreichend‘ zu bewerten ist). Darüber hinaus ergeben sich Einschränkungen und Vorgaben, etwa basierend auf dem Nutzen, den ein Datensatz zu erfüllen hat. Des Weiteren ergeben sich durch den spezifischen Anwendungsfall die im vorherigen Kapitel bereits genannten Notwendigkeiten hinsichtlich der Datengrundlage:

- Die aufgenommenen Daten haben grundlegende Anforderungen hinsichtlich Kriterien wie dem Datenschutz zu erfüllen. Sofern vertrauliche Daten enthalten sind, ist es nötig den Datensatz sicher zu speichern, so dass er nicht für Dritte zugänglich ist. Es ist weiterhin eine korrekte und für andere Nutzer verständliche Annotation des Datensatzes notwendig, um eine sinnvolle Verwendung zu gewährleisten.
- Es gilt, sich bei der Erstellung an Best Practices zu orientieren, auch wenn diese bislang nur eingeschränkt formalisiert sind. Die Erstellung gilt es sorgfältig zu dokumentieren, um eine Reproduzierbarkeit und Nachvollziehbarkeit zu gewährleisten. Neben dem Vorgehen zur Erstellung ist auch die genutzte Hardware und gegebenenfalls verwendete Software (etwa für die automatisierte Annotation) zu dokumentieren.
- Im Fall der Erstellung eines Datensatzes für einen spezifischen Anwendungsfall wie die Identifikation ist es notwendig, sich an den Vorgaben der jeweiligen Herausforderung zu orientieren. Für die Identifikation ist es notwendig, mindestens zwei Instanzen einer jeden zu identifizierenden Entität aufzunehmen. Dabei sind diese mit der gleichen oder einer anderen Kamera, am gleichen oder anderen Ort, oder lediglich zu einem anderen Zeitpunkt oder unter einem anderen Winkel aufzunehmen.

## 4.2 Identifikationsprozess

In diesem Kapitel wird die Auswahl, Nutzung und das Training der verwendeten Modelle beschrieben. Darüber hinaus wird der Identifikationsprozess, in Abhängigkeit des Anwendungsfalls, formalisiert. Zuletzt wird auf die Evaluation der erarbeiteten Ergebnisse eingegangen, insbesondere hinsichtlich des Identifikationsprozesses und der Verlässlichkeitsquantifikation.

### 4.2.1 Erstellung, Training und Auswahl der Modelle

Basierend auf den bereits vorgestellten Modellen und Identifikationsverfahren ergibt sich für diese Arbeit folgendes Vorgehen bezüglich der Auswahl der zu untersuchenden Identifikationsmodelle:

- **State-of-the-art Modell:** Es wird ein state-of-the-art Modell genutzt, angepasst für die Aufgabe der Identifikation logistischer Entitäten. Die mit diesem Modell erreichten Ergebnisse auf logistischen Entitäten dienen dem Vergleich mit den Ergebnissen im Kontext der Identifikation von Personen. Darüber hinaus dient die Nutzung dieses Modells als Bestätigung für die Übertragbarkeit der Ergebnisse eines vortrainierten Identifikationsmodells auf andere Anwendungsfälle.
- **Graphmodell:** Es wird ein eigens entwickeltes Modell für die Identifikation logistischer Entitäten verwendet. Das Modell ist unter eingeschränkter Hardwareressourcen und in Echtzeit anwendbar umzusetzen. Aus diesem Grund wird eine Graph Neural Network Architektur genutzt.
- **Ensemblemodell:** Aufbauend auf der Erstellung eines eigenen Modells und den bisher beobachteten Einschränkungen in der Identifikationsgenauigkeit von Modellen, die unter eingeschränkten Hardwareressourcen verwendbar sind (siehe Kapitel 3), wird ein Ensemblemodell verwendet. Die Nutzung eines Ensembles hat zum Ziel, die Vorteile simpler Modelle zu erhalten und gleichzeitig eine erhöhte Performance zu ermöglichen.
- **Detektionsmodell:** In einer industriellen Umgebung gilt es zuerst geeignete Bilddaten für die Identifikationsmodelle aus einer dynamischen Bewegung zu extrahieren. Es werden deshalb Detektionsmodelle für die Extraktion der nötigen Bilder für die praktische Anwendung der Identifikationsmodelle in einem industriellen Umfeld trainiert. Es ist notwendig, ein solches Detektionsmodell ebenfalls mit eingeschränkter Hardware und in Echtzeit (online) nutzbar zu entwickeln, da es bei einer on-board Nutzung einen Teil des Navigationsverfahrens des autonomen Identifikationsroboters darstellt. Die Ausgabe der Detektion gilt es gleichzeitig zur Steuerung innerhalb eines Regelkreises zu integrieren.

Basierend auf diesen Randbedingungen, wurde folgende Auswahl getroffen: Als state-of-the-art Modell wurde das in Kapitel 3 vorgestellte PCB Modell ausgewählt. Das Modell wird weiterhin in diversen Forschungsarbei-

ten verwendet und erzielt vielversprechende Ergebnisse in der Anwendung auf Menschen. Spezifisch wird die Version PCB\_P4 genutzt, welche die zu untersuchenden Bilder in vier Abschnitte unterteilt (siehe Abb. 3.1 für eine Erläuterung der Funktionsweise). Das Modell wird über 60 Epochen mit der Cross Entropy Loss Funktion trainiert. Als Optimizer wird Stochastic Gradient Descent (SGD) verwendet, die Lernrate sinkend auf 0,1 bis 0,01 eingestellt und eine Batchgröße von 20 genutzt. Eine Übersicht der Modellarchitektur ist Tab. 4.1 zu entnehmen. Für weitere Details zum Modell wird auf die einschlägige Publikation verwiesen [Sun+18].

**Tabelle 4.1** PCB Architektur unter Nutzung von ResNet-50 als Backbone Netzwerk nach [He+16; Sun+18].

Modell	Layerart	Outputformat	Aktivierung
ResNet-50	Convolutional (7 × 7)	112 × 112 × 64	ReLU
	Max Pooling (3 × 3)	56 × 56 × 64	-
	Residual Block 1	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3, 56 \times 56 \times 256$	ReLU
	Residual Block 2	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4, 28 \times 28 \times 512$	ReLU
	Residual Block 3	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6, 14 \times 14 \times 1024$	ReLU
	Residual Block 4	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3, 7 \times 7 \times 2048$	ReLU
	Average Pooling (7 × 7)	1 × 1 × 2048	-
PCB	Convolution (1 × 1)	256	ReLU
	Fully Connected	256	Softmax

Als Modell, das unter eingeschränkten Hardwareressourcen nutzbar ist, wird eine Eigenentwicklung (Graphmodell) als Lösung gewählt. Dieses Modell basiert auf der Detektion von Anomalien, die als Features zu betrachten sind. Bei einer akkuraten Auswahl der Anomalien und somit der relevanten

Features eines Bildes ist der restliche Bildinhalt für eine präzise Identifikation vernachlässigbar. Ein Großteil des Inhalts eines jeweiligen Bildes ist somit zu vernachlässigen, was den Vorteil eines solchen Ansatzes darstellt. Um dies technisch umzusetzen, werden die genutzten Anomalien in Form eines Repräsentationsgraphen dargestellt. Aus diesem Grund ist es sinnvoll, die detektierten Anomalien als Knoten des Graphen zu definieren. Es wird dabei über jeden  $16 \times 16$  px großen Bereich (mit einer 50 % Überlappung zwischen aufeinander folgenden Bereichen) iteriert und ein jeweiliger Anomaliewert berechnet, der die Wahrscheinlichkeit der Relevanz des Bildabschnitts für die Identifikation repräsentiert. Für die Berechnung des Werts wird ein simpler Anomaliedetektor verwendet. Dieser Anomaliedetektor betrachtet die beschriebenen Teilbereiche  $S$  als umso anormaler, je höher das Verhältnis der Standardabweichung von  $S$  zum Mittelwert von  $S$  ist (hier spezifisch Standardabweichung und Mittelwert der Farbe). Dieses Verhältnis  $A(S)$  ist definiert als:

$$A(S) = \frac{std(S)}{mean(S)} \quad (4.1)$$

Unter Nutzung einer Funktion wie der in Gleichung 4.1 skizzierten, ist es möglich, fälschlicherweise mehrere nebeneinanderliegende Knoten des Repräsentationsgraphen zu kreieren, die jedoch zu einem einzigen Knoten zusammengeführt werden. Um das Auftreten dieses Phänomens zu vermeiden, werden alle Teilbereiche des Bildes in absteigender Reihenfolge von  $A(S)$  durchlaufen. Wenn der aktuelle Teilbereich  $S$  nahe an einem anderen Knoten liegt, wird der Durchschnitt der Position beider Knoten berechnet (sofern die mittlere quadratische Abweichung der Position der jeweiligen Knoten über einem Schwellwert  $\tau$  liegt). Anschließend werden beide Knoten zu einem einzigen zusammengeführt. Wird der Schwellwert  $\tau$  unterschritten, so wird ein neuer Knoten erstellt. Dieser Vorgang wird wiederholt, bis insgesamt 100 Knoten generiert worden sind, die das gesamte Bild in einem stark kondensierten Repräsentationsgraphen darstellen. In jedem Knoten werden somit Informationen über eine gegebene Anomalie und ihre Umgebung gespeichert. Der zugrundeliegende Knotenfindungsalgorithmus ist in Alg. 2 beschrieben. Eine Liste aller berücksichtigten Parameter ist Tab. 4.2 zu entnehmen.

**Tabelle 4.2** Übersicht der Features, die einen Knoten des Repräsentationsgraphen ausmachen.

Position [px]	Farbe [RGB]	Dimensionalität
X-Achsen Position	Mittlere Farbe der Subregion	Anzahl an Subregionen $S_i$
Y-Achsen Position	Farbstandardabweichung	Größe der Subregion

Anschließend werden mithilfe eines Top-k-Algorithmus Graphverbindungen, basierend auf der Position der jeweiligen Anomalie generiert. Jeder Knoten wird dabei mit den fünf nächsten Knoten verbunden. Ziel dieser Methode ist es, eine minimalistische Version der Oberflächenstruktur als Graphdarstellung zu erfassen. Die Architektur des dabei entstehenden siamesischen Repräsentationsgraphmodells (Siamese Graph Neural Network) ist Tab. 4.3 zu entnehmen. Das Modell wird unter Verwendung einer Triplet Loss Funktion in Batches von 100 Bildern für bis zu 100 Epochen mit einer Patience von 10 Epochen trainiert. Es wird der Adam Optimizer und eine Lernrate von 0,001 verwendet. Das Format jeder Graphschicht wird als  $Knoten \times Merkmale + Knoten \times Knoten$  (Knotenmerkmale und Adjacency Matrix) definiert.

---

**Algorithmus 2** IRAG (Image Representation through Anomaly Graphs).
 

---

**Input:**  $S_j, pos_j$  (Subregionen und Positionen),  $nodec$  (Anzahl zu findender Knoten),  $\tau$  (Mindestabstand zwischen zwei Knoten)  
**Output:**  $nodes$  (Position der gefundenen Knoten)

```

 $n \leftarrow 0$ 
 $nodes_i \leftarrow (0,0) \forall i < nodec$ 
 $regions_i \leftarrow 0 \forall i < nodec$ 
 $a_i \leftarrow A(S_i) \forall S_i$ 
 $dex \leftarrow reverse(argsort(a_i))$ 
 $k \leftarrow 0$ 
while  $n < nodec$  do ▷ Loop für 100 Knoten
   $curr \leftarrow dex[k]$ 
   $dist \leftarrow mse(pos_{curr} - nodes[j]) \forall j < n$ 
  if  $min(dist) < \tau$  then ▷ Überprüfung ob Schwellwert unterschritten wird
     $closest = argmin(dist)$ 
     $nodes_{closest} \leftarrow \frac{nodes_{closest} \cdot regions_{closest} + pos_{curr}}{regions_{closest} + 1}$ 
     $regions_{closest} \leftarrow regions_{closest} + 1$ 
  else ▷ Falls Schwellwert überschritten wird
     $nodes_n \leftarrow pos_{curr}$ 
     $regions_n \leftarrow 1$ 
     $n \leftarrow n + 1$ 
  end if
   $k \leftarrow k + 1$ 
end while
return  $nodes$ 

```

---

Als Ensemblemodell wurde eine Kombination von vier simplen Modellen gewählt. Alle vier Modelle verwenden den Ansatz eines Siamese Neural Network, sind aber jeweils auf andere Charakteristika der Bilddaten fokussiert:

- CNN-Submodell: Es wird ein bildbasiertes Modell (CNN) verwendet, das auf konventionelle Weise Informationen aus dem gesamten Bild extrahiert. Dies stellt jedoch auch den zeitaufwändigsten Ansatz dar und ist am anfälligsten für Overfitting.
- Helligkeits-Submodell: Es wird ein helligkeitsbasiertes Modell verwendet, bei dem jeweils der Grayscale-Mittelwert pro  $16 \times 16$  px Teilabschnitt eines Bildes als Identifikationsmerkmal verwendet wird.
- Farb-Submodelle: Zwei weitere Modelle beschäftigen sich mit dem mittleren RGB-Wert und der Farbvarianz, ebenfalls  $16 \times 16$  px großer Teilabschnitte. Es wird dementsprechend jeweils der Mittelwert jeder Farbe und die Standardabweichung jeder Farbe eines jeden Teilabschnitts bestimmt.

Als weiteres Modell dient das bereits beschriebene Repräsentationsgraphmodell. Es wird jeweils eine 50 % Überlappung der Teilabschnitte verwendet und 1 oder 3 Werte pro Teilabschnitt berechnet (je nachdem, ob Grayscale oder RGB-Farben verwendet werden), was zu einem 768 beziehungsweise 2.304 dimensionalen Vektor führt. Jedes dieser Submodelle des Ensembles verwendet drei Dense Layer mit jeweils 100 Knoten, um eine 100-dimensionale Darstellung zu generieren.

**Tabelle 4.3** Architektur des Siamesischen Graph Neural Network.

Layerart	Outputformat	Aktivierung
Input	$50 \times 10 + 50 \times 50$	-
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Graph Convolution	$50 \times 256 + 50 \times 50$	SELU
Global Mean Pooling	256	-
Fully Connected	100	ReLU
Fully Connected	100	ReLU
Fully Connected	100	Linear

Die verwendeten Modelle werden mittels der Ensemblingmethode der Konkatenation kombiniert. Die Konkatenation ist eine Ensemblingmethode, bei der die einzelnen Outputs der Modelle in einer Transformation  $T$  an-

einandergereiht werden. Diese Transformation normalisiert jedes Ensemble-Element unter Verwendung des Mittelwerts und der Standardabweichung und reiht sie aneinander. Sie ist als eine Ensemblingmethode zu betrachten, die dem Konzept des Bagging folgt:

$$T_{\text{Concat}}(\{x_0^i, \dots, x_n^i\}) = \{z_0^i, \dots, z_n^i\} \quad \text{mit} \quad z_j^i = \frac{x_j^i - \bar{x}_j}{s_{x_j}} \quad (4.2)$$

Hier bezeichnet  $x_j^i$  die  $j$ -te Komponente der Darstellung des Bilds  $x^i$ ,  $\bar{x}_j$  ihren Mittelwert und  $s_{x_j}$  ihre Standardabweichung. Alle drei Modelle werden pro Datensatz fünffach kreuzvalidiert trainiert und getestet.

Als vorgeschaltetes Detektionsmodell wurde YOLO [Red+16] ausgewählt. Spezifisch wird Tiny YOLO Version 4 gewählt, da es sich angesichts der genannten Anforderungen und Hardwareeinschränkungen besonders für den Anwendungsfall eignet. Das Modell in seiner Standardkonfiguration wird für 1.000 Epochen mit einer Batchgröße von 32 trainiert.

## 4.2.2 Beschreibung des Identifikationsprozesses

Unabhängig von der Nutzung eines spezifischen Modells oder Datensatzes gilt es den Prozess der Identifikation zu beschreiben, wie er in dieser Arbeit verstanden wird und fortan genutzt wird. Basierend auf den bereits eingeführten Definitionen aus Kapitel 2, werden die Begriffe Query und Gallery wie folgt formal definiert:

### **Definition:** *Query & Gallery*

Unter Annahme eines Query Sets  $\mathcal{Q} \subset \mathcal{D}$  und eines Gallery Sets  $\mathcal{G} \subset \mathcal{D}$  mit  $\mathcal{Q} \cap \mathcal{G} = \emptyset$  ist das Problem der Wiedererkennung eines Bildes  $x_i \in \mathcal{Q}$  einer ID  $i$  beschreibbar als die Aufgabe ein Bild  $y_j \in \mathcal{G}$  der ID  $j$  zu finden, sodass die IDs übereinstimmen.

Aufbauend auf dieser Zuordnung an Bildern aus Query und Gallery sieht der Prozess der Identifikation folgende Schritte vor:

- Die Nutzung eines Bildes als Input für das Identifikationsmodell
- Die Extraktion der relevanten Features des Bildes durch das Modell
- Der Vergleich des Bildes (Query) mit den bereits vorhandenen Bildern der Datenbank (Gallery)
- Die Zuordnung einer ID, basierend auf dem vorangegangenen Vergleich, sofern ein festgelegter Neuheitsschwellwert nicht überschritten wird
- Das Hinzufügen einer neuen ID, sofern der Neuheitsschwellwert überschritten worden ist

Für den Fall eines Einsatzes auf Daten aus einem Videostream, anstelle eines Datensatzes (einer Online Evaluation), gehen den oben beschriebenen Schritten weitere voraus. So wird zunächst ein Detektionsmodell verwendet, um die gewünschten Objekte aus einem Videostream zu extrahieren. Basierend auf den Bounding Boxes der Detektion werden Standbilder des Videostreams zugeschnitten und die resultierenden Bilder als Input für das Identifikationsmodell genutzt. Da sich eine solche Online Evaluation deutlich in ihrem Vorgehen von einer Offline Evaluation (der Evaluation unter Nutzung eines kuratierten Datensatzes) unterscheidet, werden die beiden Vorgehen fortan als Szenario I (Offline Evaluation) und Szenario II (Online Evaluation) bezeichnet und getrennt betrachtet. Szenario I wird sich insofern mit der detaillierten Analyse der bereits beschriebenen Modelle auf geeignete Datensätze befassen. In Szenario II wird das Identifikationsverfahren in einer industrieähnlichen Umgebung untersucht und mittels eines AGV live auf eine Auswahl logistischer Entitäten (Euro-Paletten) angewandt.

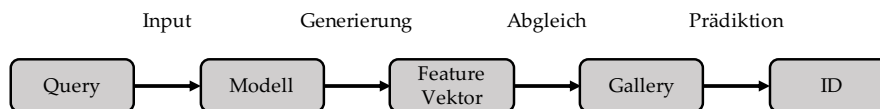


Abb. 4.1 Allgemeiner Workflow der Identifikation in dieser Arbeit.

Nach dem Abgleich der Bilder aus Query und Gallery findet weitergehend eine Auswertung der Ergebnisse (True Positive, False Positive, True Negative, False Negative, siehe Kapitel 2.2) statt. Der übergreifende Prozess der Identifikation in Form eines Flussdiagramms ist Abb. 4.1 zu entnehmen.

### 4.2.3 Auswahl der Evaluationsmetriken

Zur Evaluation der Versuchsergebnisse, die sich aus den beiden beschriebenen Szenarien ergeben, werden folgende Evaluationsmetriken verwendet:

- **Ranked Accuracy (Rank-k-Accuracy):** Die prozentuale Genauigkeit der Zuordnung der IDs der verwendeten Bilder, sei es im Hold-out Datensatz in Szenario I oder der zu identifizierenden Palettenklötze in Szenario II. Die Ergebnisse werden nach geringsten Distanzen im Darstellungsraum, beziehungsweise der größten Ähnlichkeit bis zum fünften Rang sortiert (Rank-5-Accuracy). Die Kalkulation der Distanzen basiert auf den Auswertungen der jeweiligen Modelle.
- **Neuheitsschwellwert:** Für Szenario II wird ein Schwellwert einer Distanz von 1,5 definiert, der als Entscheidungskriterium für die Serialisierung oder Identifikation einer ID dient. Bei Überschreiten dieses Schwellwerts wird eine neue ID in der Galerie angelegt. Der Ansatz ist angelehnt an die Vorgehensweise, die in einer Vorarbeit erarbeitet wurde [Sch+23] und ist empirischer Natur.

- Intersection over Union (*IoU*): Im Fall von Szenario II wird neben der Ranked Accuracy auch die Güte der Detektion untersucht. Die YOLO-basierte Detektion wird mittels *IoU* beim Trainingsprozess der Modelle evaluiert. Zur Detektion wird ein Schwellwert von einer Distanz von 1,5 gesetzt, ab welchem von einer erfolgreichen Detektion ausgegangen wird und ein Bild erstellt wird, das für die Identifikation genutzt werden.

Eine Übersicht der verwendeten Evaluationsmetriken sowie ihrer Zusammenhänge ist Abb. 4.2 zu entnehmen.

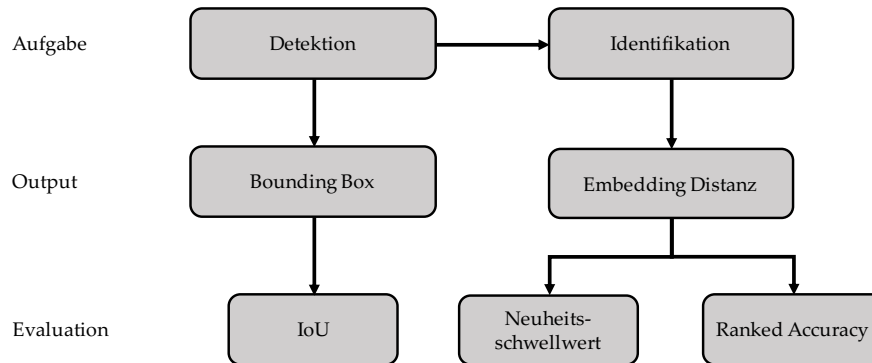


Abb. 4.2 Workflow der Metriken für Szenario II.

### 4.3 Verlässlichkeitsbetrachtung

In diesem Abschnitt wird die Verlässlichkeit sowohl des Identifikationsprozesses als auch von Datensätzen und Modellen des Maschinellen Lernens betrachtet. Dafür werden in einem ersten Schritt die Schlüsselaspekte der Verlässlichkeit hergeleitet. Diese werden jeweils aus einer Literaturrecherche deduziert und anschließend tabellarisch zusammengefasst. Für die Recherche der relevanten Aspekte der Verlässlichkeit wurde im Rahmen einer Vorarbeit [Rut+24b] die Suchmaschine „Google Scholar“ verwendet. Es wurden einschlägige Datenbanken durchsucht und relevante Literatur zum Thema der Verlässlichkeit im Maschinellen Lernen gesichtet. Insbesondere über Review Artikel wurden die Schlüsselbegriffe in diesem Forschungsfeld identifiziert. Es wurde nach Definitionen und Metriken ebenjener Begriffe und dem Begriff der Verlässlichkeit selbst gesucht. Basierend auf dieser Recherche werden in diesem Kapitel jeweils operationelle Definitionen vorgeschlagen, die einem ersten Schritt zur Quantifikation der Begrifflichkeiten dienen. Des Weiteren wird die in dieser Arbeit verwendete Herangehensweise zur Evaluation des Konzepts der Verlässlichkeit beschrieben.

### 4.3.1 Herleitung der relevanten Aspekte der Verlässlichkeit

Im Kontext der Bewertung von Modellen des Maschinellen Lernens und ihren zugehörigen Datensätzen ist das Konzept der Fairness eine häufig anzutreffende Anforderung. Der Begriff ist jedoch aktuell nicht standardisiert definiert und stattdessen durch mehrere, teils unvereinbare, Umschreibungen repräsentiert [Win21]. Fairness im Kontext der Entscheidungsfindung ist ein grundlegender Baustein, der auf den Prinzipien der gleichen Behandlung und der Vermeidung von Ungleichheiten und Diskriminierung beruht [Dod+19]. Technischer beschrieben, wird Fairness als der Versuch verstanden, unvoreingenommene Ground-Truth Werte für das Training von Modellen zu verwenden [Sch+22]. Dennoch gilt es als herausfordernd, Fairness zu messen [Tor+20], insbesondere im Bereich des Maschinellen Lernens [CH24]. Einige Ansätze für Fairness-Metriken wurden bereits vorgeschlagen [Sch+22]. Allerdings handelt es sich dabei um solche, die spezifisch für ihren respektiven Anwendungsfall nutzbar sind und somit keine Generalisierung ermöglichen. Dies liegt an der Abhängigkeit der Bedeutung des Konzepts der Fairness vom jeweiligen Kontext, in welchem es verwendet wird: Während Fairness möglicherweise nicht relevant für Modelle oder Daten ist, die in Produktionsanlagen verwendet werden, ist sie für Anwendungen, die kritische, personenbezogene Entscheidungen unterstützen entscheidend, wie etwa eine Kreditvergabe oder die Evaluation von Bewerbern [Sch+22].

Fairness wird in manchen Kontexten weiterhin in individuelle und Gruppenfairness unterteilt. Bei individueller Fairness geht es darum, gerechte Ergebnisse für Personen mit ähnlichen Merkmalen sicherzustellen [Kus+17; Zem+13]. Ein anderer Ansatz in diesem Kontext besteht darin, vollständig unveränderliche Attribute als Beurteilungskriterien auszuschließen, die demnach zu Diskriminierung führen würden [Thu22]. Das gegenteilige Konzept, die sogenannte kontrafaktische Fairness, wird jedoch von anderen Forschern ebenfalls als Fairnessgarant vorgeschlagen [Kus+17]. Gruppenfairness betont die Bedeutung der Verminderung von mit bestimmten Attributen verbundenen Voreingenommenheiten [CH24; Dod+19; SS22] und stellt demnach die Gleichbehandlung von Untergruppen gegenüber Mehrheitsgruppen sicher [HPS16]. Die relevanten, oft demografischen Gruppenattribute [CH24; Kus+17], umfassen Faktoren wie Ethnizität und Geschlecht. [SS22] warnt davor, der Fokus auf Gruppenfairness würde zu Kosten der Privatsphäre von Individuen zu garantieren sein, da sensible personenbezogene Daten und Attribute zu diesem Zweck offengelegt werden würden. Die relative identische Verteilung der Ergebnisse über alle Kategorien hinweg sei dabei nicht erforderlich. Vielmehr sei hierbei das Ziel, die relative Verteilung der erhaltenen Ergebnisse sei mit der Verteilung der Attribute im ursprünglichen Datensatz in Übereinstimmung zu bringen [Zem+13].

**Tabelle 4.4** Überblick der relevanten Literatur zu Fairness im Maschinellen Lernen.

Referenz	Jahr	Autor(en)	Verständnis des Begriffs der Fairness	Met.	Def.
[Sch+22]	2022	Schmitz et al.	Nutzung unvoreingenommener Ground-Truth Werte	✓	✗
[Thu22]	2022	Thuraisingham	Unvoreingenommene Entscheidungen, Ausschluss geschützter Entscheidungskriterien	✗	✗
[SS22]	2022	Strobel et al.	Ungleiche Behandlung demographischer Gruppen aufgrund ungleicher Trainingsdaten	✗	✗
[TLS21]	2021	Thiebes et al.	Gleichmäßige Verteilung von Vorteilen, Minimierung von Ungleichheit und Leid	✗	✓
[Tor+20]	2020	Toreini et al.	Bereitstellung von nicht-diskriminierenden und unvoreingenommenen Ergebnissen	✗	✗
[CH24]	2020	Caton et al.	Parität in Ergebnisverteilungen, proportionale Zuschreibungen, Behandlungsgleichheit	✓	✓
[Dod+19]	2019	Dodge et al.	Kein Unterschied in der Behandlung von Personen, die sich nur in einem geschützten Merkmal unterscheiden	✗	✗
[Kus+17]	2017	Kusner et al.	Entscheidungen bleiben sowohl im realen Szenario als auch in einem kontrafaktischen Szenario gleich, in welchem die demografische Gruppe einer Person gewechselt wird	✓	✓
[HPS16]	2016	Hardt et al.	Geschützte Attribute sind bei Entscheidungsfindungen nicht zu berücksichtigen	✓	✓
[Zem+13]	2013	Zemel et al.	Ausschluss von Informationen, welche die Zugehörigkeit einer Person zu einer Gruppe implizieren (Fairness durch Unwissenheit)	✗	✓
[Mit80]	1980	Mitchell	Älteste identifizierte Publikation zu Biases im Maschinellen Lernen	✗	✗

Basierend auf der durchgeführten Recherche zum Begriff der Fairness, die in Tab. 4.4 in ihren relevantesten Publikationen zusammengefasst ist, wird zum Zwecke der Operationalisierung folgende Definition vorgeschlagen:

**Definition:** *Fairness*

Fairness im Maschinellen Lernen gewährleistet die gleiche Behandlung sowohl von Gruppen als auch von Individuen und verhindert Voreingenommenheiten, um gerechte Ergebnisse zu fördern.

Ein weiterer relevanter Begriff im Kontext der Verlässlichkeit ist die Robustheit. Das Konzept der Robustheit befasst sich mit der Gewährleistung erwarteter Modellleistungen, selbst unter intensiver Prüfung oder bei Auftreten von Randphänomenen, wie außergewöhnlichen, manipulierten, veräuschten oder unbekanntem Daten [Dre+21; GMP18; Li+23; SS22]. [Dre+21] verstehen Robustheit als ‚überladener‘ Begriff, der eine Vielzahl von Interpretationen zuließe. Die Autoren geben an, 53 Studien analysiert zu haben, die nicht definieren, was sie unter Robustheit verstehen, den Begriff jedoch für ihre Untersuchungen nutzen. Daher variieren die Interpretationen der jeweiligen Forscher von bloßer Task Performance auf hold-out Datensätze bis hin zur Aufrechterhaltung dieser Performance auf manipulierten Eingaben (beispielsweise zusätzlichem Rauschen oder adversarial attacks). Die Änderung der Leistung bei Randfällen, Ausnahmen oder beschädigten Daten ist ebenfalls als Teil der Robustheit zu betrachten.

Laut [SS22] besteht Robustheit vor allem darin, sich gegen adversarial attacks zu wehren. Ein robustes Modell sei nicht nur in der Lage, das Label für gewöhnliche, erwartete Inputdaten korrekt vorherzusagen, sondern auch dann richtige Prädiktionen auszugeben, wenn die Inputdaten manipuliert worden sind [Li+23]. In diesem Sinne sei der Output eines robusten Modells nicht übermäßig empfindlich hinsichtlich der Änderung des Inputs [Win21]. Darüber hinaus erwähnen die Autoren das Risiko von crowd-sourced Daten. Die Qualität ebenjener Daten würde aufgrund des Datenerfassungsprozesses veräuscht sein. Ebenfalls würde das unpräzise oder gar vorsätzlich böswillige Handeln der Crowd Worker nicht auszuschließen sein. [SS22] konstatieren, die meisten modernen Modelle des Maschinellen Lernens seien nicht robust gegenüber solchen Gefahren. Zur Minderung dieser Risiken wurden Strategien vorgeschlagen, wie die Reduzierung der Sensitivität eines Modells gegenüber Änderungen in Teilen der Trainingsdaten oder der Ansatz, Ausreißer während des Trainings vollständig zu ignorieren [SS22]. [Xu+21] schlagen in diesem Kontext die Verwendung von Graphmodellen als besonders robuste Modelltypen vor. [DCH21] betonen abermals die Bedeutung der Robustheit eines Modells gegenüber Perturbationen und Distribution Shifts.

**Tabelle 4.5** Überblick der relevanten Literatur zu Robustheit im Maschinellen Lernen.

Referenz	Jahr	Autor(en)	Verständnis des Begriffs der Robustheit	Met.	Def.
[Li+23]	2023	Li et al.	Die Fähigkeit, mit Ausführungsfehlern, fehlerhaften Inputs oder unbekanntem Daten umzugehen	✓	✓
[SS22]	2022	Strobel et al.	Korrekte Prädiktionen für gewöhnliche Eingaben als auch manipulierte oder veräuschte Daten	✗	✗
[Win21]	2021	Wing	Wie sensibel der Output des Systems auf eine Änderung des Inputs reagiert	✗	✓
[Dre+21]	2021	Drenkow et al.	Die Gewährleistung der erwarteten Performance bei manipulierten oder modifizierten Daten sowie die Generalisierbarkeit über verschiedene Domänen hinweg	✓	✓
[Xu+21]	2021	Xu et al.	Gewährleistung desselben Outcomes trotz veräuschter Daten oder adversarial attacks	✓	✗
[DCH21]	2021	Darestani et al.	Sicherstellung korrekter Prädiktionen trotz Perturbationen oder Distribution Shifts	✓	✗
[GMP18]	2018	Goodfellow et al.	Akkurate Prädiktionen auch für Anomalien und für modifizierte Inputs bereitzustellen	✗	✗
[RBB17]	2017	Rauber et al.	Die Widerstandsfähigkeit eines Modells gegenüber Perturbationen der Daten	✓	✓
[Ash76]	1976	Ashton	Älteste identifizierte Publikation zu Modellrobustheit	✗	✗

Zusammenfassend lässt der Mangel an einheitlichen Definitionen und die Vielfalt der Umschreibungen des Begriffs kaum eine Quantifizierung oder Vergleichbarkeit der Robustheit im Kontext des Maschinellen Lernens zu. Basierend auf der durchgeführten Recherche zum Begriff der Robustheit, die in Tab. 4.5 in ihren relevantesten Publikationen zusammengefasst ist, wird zum Zwecke der Operationalisierung folgende Definition vorgeschlagen:

**Definition:** *Robustheit*

Robustheit im Maschinellen Lernen bezeichnet die Fähigkeit, eine erwartete Leistung trotz außergewöhnlicher, manipulierter oder veräuschter Daten aufrechtzuerhalten.

Ein mit der Robustheit verwandter Begriff ist die Integrität. Integrität beschreibt verschiedene Konzepte im Bereich des Maschinellen Lernens, wie beispielsweise die Genauigkeit der verwendeten Daten oder des Modells

sowie die prozedurale Unversehrtheit [Thu22]. Einflüsse wie die Korruption von Daten oder die unerwünschte Änderung der Modellarchitektur haben ebenfalls Auswirkungen auf die Integrität [Thu22]. [HZL18] betonen die Bedeutung der Modellintegrität, insbesondere bei einer remote Bereitstellung, bei welcher möglicherweise Änderungen am Modell vorgenommen werden, ohne dass autorisierte Benutzer diese Änderungen validieren oder gar bemerken. Analog empfehlen [TLS21] die Verwendung der Distributed Ledger Technologie, um sowohl unerwünschte Änderungen in einem System zu verhindern als auch die Rückverfolgbarkeit der vorgenommenen Änderungen sicherzustellen. Neben der Verhinderung solcher Manipulationsversuche nennen [Tor+20] die rechtzeitige Erkennung und anschließende Reparatur manipulierter Daten als kritische Maßnahme zur Gewährleistung der Integrität von Modellen und Daten im Maschinellen Lernen. Daher besteht die Gewährleistung der Datenintegrität darin, Änderungen wahrzunehmen und diese entsprechend nachzuverfolgen [MB20].

**Tabelle 4.6** Überblick der relevanten Literatur zu Integrität im Maschinellen Lernen.

Referenz	Jahr	Autor(en)	Verständnis des Begriffs der Integrität	Met.	Def.
[Thu22]	2022	Thuraisingham	Die Genauigkeit der verwendeten Daten, die prozedurale Kohärenz und die algorithmische Korrektheit	✗	✗
[Hou+22a]	2022	Hou et al.	Die Verhinderung böswilliger Änderungen der Trainingsdaten, die darauf abzielen, die Prädiktionen des Modells zu manipulieren	✗	✗
[Kut+22]	2022	Kuttichira et al.	Die Verhinderung von Modellmanipulation durch den Vergleich von Prädiktionen des originalen und des potenziell beeinträchtigten Modells	✓	✗
[Tor+20]	2020	Toreini et al.	Die Verhinderung von Manipulationen sowie die Detektion und Wiederherstellung manipulierter Daten	✗	✗
[MB20]	2020	Meske et al.	Änderungen an den Daten erfolgen nicht unbemerkt und sind nachvollziehbar	✗	✗
[HZL18]	2018	He et al.	Die Sicherstellung, der Unversehrtheit der Architektur des Modells, wenn es remote bereitgestellt wird	✓	✗
[HS91]	1991	Haber et al.	Älteste identifizierte Publikation zu Datenintegrität	✗	✗

Traditionell verwenden Integritätsnachweisverfahren eine Art von Hashingverfahren. Solche Hashingverfahren sind jedoch im Fall des Maschinellen Lernens kaum anwendbar, da der Zugang zum verwendeten Modell

oft eingeschränkt ist [Kut+22]. [Kut+22] schlagen daher einen Bayesian Compromise Detection Algorithmus vor, der darauf abzielt, den Unterschied in der Vorhersage eines Originalmodells zu einem potenziell kompromittierten Modell zu vergleichen. Die Autoren schlagen somit ein impliziertes Maß für die Integrität eines Modells vor: die Ähnlichkeit des kompromittierten zum ursprünglichen Modell. Eine formale Definition des Begriffs der Integrität bieten die Autoren allerdings nicht an. Analog zu diesem Ansatz schlagen [Hou+22a] eine auf Ähnlichkeit basierende Integritätsschutzmethode für Deep Learning Systeme vor. Diese Methode basiert auf einem Anomalieerkennungansatz, der die Ähnlichkeit zwischen verdächtigen Datensamples und Samples in einem Verifikationsdatensatz misst. Auch in diesem Fall ermöglicht dies die Gewährleistung der Integrität und stellt ein impliziertes Integritätsmaß dar.

Trotz der ersten Schritte hin zu einem Maß der Integrität mangelt es an übergreifenden und expliziten Quantifikationsverfahren und formalen Definitionen in diesem Kontext. Basierend auf der durchgeführten Recherche zum Begriff der Integrität, die in Tab. 4.6 in ihren relevantesten Publikationen zusammengefasst ist, wird zum Zwecke der Operationalisierung folgende Definition vorgeschlagen:

**Definition:** *Integrität*

Integrität im Maschinellen Lernen ist der Versuch, unerfasste oder unbefugte Änderungen an Daten oder Modellen zu verhindern, die Risiken von Manipulationsversuchen zu mindern und sie gegebenenfalls rückgängig zu machen.

Im Bereich des Maschinellen Lernens hat die Erklärbarkeit als entscheidender Aspekt der Verlässlichkeit an Bedeutung gewonnen, der das Vertrauen der Nutzer, ethische Überlegungen und die gesellschaftliche Akzeptanz von sogenannten KI-Technologien beeinflusst [BH21]. Dennoch ist das Konzept der Erklärbarkeit aktuell nur vage in der Forschung definiert. Forscher sind sich weitgehend bezüglich der mehrdeutigen Verwendung des Begriffs einig, welcher oft synonym mit Begriffen wie Transparenz, Verständlichkeit, Nachvollziehbarkeit und Interpretierbarkeit verwendet wird [Arr+20; BH21; Gui+18; Sch+22; Tor+20]. Es wurden bereits Taxonomien der Erklärbarkeit vorgeschlagen, etwa im Hinblick auf das Verständnis der Funktionsweise eines Modells (pre-hoc) oder lediglich seiner Entscheidungen (post-hoc) [Her+23].

**Tabelle 4.7** Überblick der relevanten Literatur zu Erklärbarkeit im Maschinellen Lernen.

Referenz	Jahr	Autor(en)	Verständnis des Begriffs der Erklärbarkeit	Met.	Def.
[Her+23]	2023	Herm et al.	Übersicht der Wie, Wieso, Wieso Nicht und Wie Sonst Ansätze der Erklärbarkeit	✗	✓
[Li+23]	2023	Li et al.	Die Fähigkeit zu verstehen, wie ein Modell seine Entscheidungen trifft	✗	✓
[Sch+22]	2022	Schmitz et al.	Technische Transparenz künstlicher Intelligenz	✗	✓
[SS22]	2022	Strobel et al.	Übereinstimmung von Modellsemantik und Aufgabensemantik	✗	✗
[BH21]	2021	Burkart et al.	Die Bereitstellung von Begründungen für getroffene Entscheidungen	✗	✗
[Win21]	2021	Wing et al.	Die Fähigkeit, die Entscheidungen des Modells auf eine für Menschen verständliche Weise zu rechtfertigen	✗	✓
[Tor+20]	2020	Toreini et al.	Die Erklärung und Interpretation von Entscheidungskonsequenzen gegenüber Stakeholdern	✗	✗
[Arr+20]	2020	Arrieta et al.	Die Bereitstellung von Modelldetails und -entscheidungen, um die unterliegende Funktionsweise transparent und leicht verständlich zu machen	✗	✓
[MV20]	2020	Marcinkevics et al.	Menschenverständliche Erklärung von der Entscheidungen von intransparenten Modellen	✓	✗
[MSM18]	2018	Montavon et al.	Übertragung abstrakter Entscheidungsfindungskonzepte in eine für Menschen verständliche Domäne	✗	✓
[Gui+18]	2018	Guidotti et al.	Der Grad der Modell-, Prädiktions- oder Datenverständlichkeit aus Sicht eines Menschen	✗	✓
[AHM95]	1995	Auer et al.	Älteste identifizierte Publikation zu erklär-baren Modellen	✗	✗

Im Rahmen der Analyse der pre-hoc Erklärbarkeit verstehen [Arr+20; SS22] Erklärbarkeit als die semantische Übereinstimmung des Modells mit der respektiven Aufgabensemantik. Im Gegensatz dazu verstehen [BH21; Li+23; Tor+20; Win21] das Konzept der Erklärbarkeit als die Bereitstellung von Gründen für die von einem Modell getroffenen Entscheidungen oder Prädiktionen. In diesem Zusammenhang betonen [Her+23] die Bedeutung der Kenntnis über die jeweiligen Nutzergruppen, da dies bestimmt, wie die Erklärbarkeit geformt wird. So wird eine Adaption der Erklärungstiefe oder -art, in Abhängigkeit der Stakeholder als relevante Anforderung der Erklärbarkeit genannt. Die Autoren führen darüber hinaus einen Versuch durch, bei welchem Nutzer die Erklärbarkeit einer Reihe von Modellen auf einer Likert-Skala bewerten. Basierend auf dieser Bewertung bieten die Autoren eine Möglichkeit zur Bewertung der Erklärbarkeit eines Modells an.

Einige Forschungsarbeiten fokussieren sich weiterhin auf die Vorstellung der Erklärbarkeit im Hinblick auf das menschliche Verständnis [Arr+20; Gui+18; MSM18; MV20; Tor+20]. Diese Aufgabe variiert je nach Modell und die Intransparenz bestimmter Modelltypen, wie beispielsweise Neural Networks im Kontrast zu Support Vector Machines, wird in diesem Zusammenhang diskutiert [MV20]. Die zunehmende Tendenz der Nutzung von Black Box Modellen, wird in diesem Kontext als zusätzliche Herausforderung der Gewährleistung der Erklärbarkeit betrachtet.

Im Kontrast zur Erklärbarkeit von Modellen wird die Erklärbarkeit von Daten in der aktuellen Forschung vernachlässigt. [Gui+18] diskutieren Daten im Kontext des menschlichen Verständnisses zusammen mit Modellen und ihren Vorhersagen. Die Autoren verstehen Bilder und Texte als erklärbarer als tabellarische, Vektor- oder Matrixdaten. Laut den Autoren bieten erstere den Nutzern einen intuitiveren Zugang zur Dateninterpretation.

Zusammenfassend zeichnen diese Studien ein umfassendes Bild der sich entwickelnden Landschaft der Erklärbarkeit im Maschinellen Lernen und betonen ihre facettenreiche Natur sowie ihre Rolle bei der Förderung des Verlässlichkeitsempfindens seitens der Nutzer. Dennoch sind keine standardisierten Metriken oder Definitionen der Erklärbarkeit vorhanden. Basierend auf der durchgeführten Recherche zum Begriff der Erklärbarkeit, die in Tab. 4.7 in ihren relevantesten Publikationen zusammengefasst ist, wird zum Zwecke der Operationalisierung folgende Definition vorgeschlagen:

**Definition:** *Erklärbarkeit*

Erklärbarkeit im Maschinellen Lernen ist der systematische Versuch, Entscheidungsfindungsverfahren von Modellen interpretierbar und Datensätze verständlich zu gestalten, um sowohl Einblicke in ihre inneren Abläufe zu ermöglichen als auch Stakeholder bei der Validierung von Ergebnissen zu unterstützen.

Zuletzt stellt auch der Begriff der Sicherheit einen wichtigen Aspekt der Verlässlichkeit im Kontext des Maschinellen Lernens dar. Der Begriff bezieht sich auf die Sicherheit von Modellen und Daten sowie den damit verbundenen Schutz der Privatsphäre, wie in Artikel 4 der Datenschutzgrundverordnung (DSGVO) definiert [Eur16]. Kürzlich hat die Europäische Union darüber hinaus den Artificial Intelligence Act veröffentlicht, der die Sicherheit als eine seiner Bedenken im Zusammenhang mit der Verwendung von KI-Lösungen umfasst [Eur23].

Der Begriff der Sicherheit wird oft synonym mit Begriffen wie Datenschutz und Zuverlässigkeit verwendet [JLE14; Moh+22; SS22; Thu22]. Im Englischen ist darüber eine signifikante Überlappung der Begriffe Safety und Security festzustellen. Ähnlich wie für den Begriff der Fairness wird während des Nutzungsprozesses zwischen Daten unterschieden, die entweder Individuen oder Personengruppen betreffen [SS22]. In diesem Zusammenhang bieten [Moh+22] eine Checkliste von Sicherheitsmaßnahmen und -strategien an, die als Teil einer Metrik geeignet sind. [SM21] fokussieren sich auf den Begriff der Privatsphäre und die damit verbundenen Datenschutzrisiken bei der Verwendung von geschützten Daten im Maschinellen Lernen. Die Autoren nennen ‚membership inference attacks‘ als eine Bedrohung der Privatsphäre, da diese Angriffe versuchen festzustellen, ob ein bestimmter Datenpunkt zum Training des Modells verwendet wurde oder nicht. Dies führt zu der Offenlegung vertraulicher und sensibler Informationen über Individuen, etwa wenn die Gruppenzugehörigkeit einer Person zu einer Gruppe bekannt wird, die in einem gesundheitsbezogenen Datensatz erfasst ist. Die betrifft beispielsweise Personen, die an einer Krankheit leiden, die es geheimzuhalten gilt. Dieses Phänomen der indirekten Ableitung vertraulicher Informationen wurde bereits 1977 beschrieben und als statistische Offenlegung benannt [Dal77].

Analog zu dieser Herausforderung führen [De 20] das Konzept der differentiellen Privatsphäre ein. Dieses Phänomen befasst sich mit dem Privatsphärenrisiko, impliziert durch die Unterschiede in der Ausgabe eines Modells, wenn ein einzelner Datenpunkt entfernt wird, womit potenziell vertrauliche Informationen über eine Person preisgegeben werden. Andere Forschungsarbeiten fokussieren sich auf die Vertraulichkeit der Modelle selbst (also die Architektur des Modells und spezifische Parameter), die mög-

licherweise urheberrechtlich geschützt und daher vertraulich sind [Pap+18]. Die Offenlegung eines solchen intellektuellen Eigentums und möglicherweise unrechtmäßige Umgehung von Patentrechten stellt ebenfalls eine Sicherheitsbedrohung dar. Darüber hinaus, beeinträchtigt dies potentiell auch die Privatsphäre der verwendeten Datenquelle, insbesondere wenn die Nutzer des Modells sich nicht für die Nutzung autorisieren (etwa im Fall von Modellen, die öffentlich verfügbar sind).

**Tabelle 4.8** Überblick der relevanten Literatur zu Sicherheit im Maschinellen Lernen.

Referenz	Jahr	Autor(en)	Verständnis des Begriffs der Sicherheit	Met.	Def.
[Moh+22]	2022	Mohseni et al.	Ein Konzept zum Schutz vor unerwünschten Ereignissen, wie Datendiebstahl und Datenschutzverletzungen	✗	✓
[Sch+22]	2022	Schmitz et al.	Die Sicherstellung der gewünschten Funktionsweise eines Systems, das vor potentiellen Schwachstellen geschützt wird	✗	✗
[SS22]	2022	Strobel et al.	Die sichere Verwahrung von Datensatzinformationen	✗	✓
[Thu22]	2022	Thuraisingham	Modelle greifen nur auf die Daten zu, für die sie autorisiert sind, um die zugewiesene Aufgabe auszuführen; Daten gewährleisten hohe Prädiktionsgenauigkeit bei gleichzeitiger Privatsphäre- und Datenschutzverwahrung	✗	✓
[SM21]	2021	Song et al.	Sicherstellung der Anonymität von Personen, deren persönliche Informationen Teil von Datensätzen sind	✗	✗
[De 20]	2020	Decristofaro	Die Sicherstellung, der niedrigen Abweichung der Outputs eines Modells für zwei Datensatzversionen, die sich nur durch einen einzelnen Datenpunkt unterscheiden	✗	✓
[Pap+18]	2018	Papernot et al.	Die Vertraulichkeit der Architektur und Parameter eines Modells sowie der Datenquellen	✗	✗
[JLE14]	2014	Zhanglong et al.	Der Schutz privater Informationen vor Datenlecks, insbesondere von Informationen über Einzelpersonen	✗	✓
[Dal77]	1977	Dalenius	Älteste gefundene Publikation zu Datensatzsicherheit	✗	✗

Auch für den Begriff und das Konzept der Sicherheit existieren somit bislang noch keine einheitlichen Definitionen oder Metriken, oder gar ein gemeinsames Verständnis für die Begrifflichkeit selbst. Basierend auf der durchgeführten Recherche zum Begriff der Sicherheit, die in Tab. 4.8 in ihren relevantesten Publikationen zusammengefasst ist, wird zum Zwecke der Operationalisierung folgende Definition vorgeschlagen:

**Definition:** *Sicherheit*

Sicherheit im Maschinellen Lernen umfasst den Schutz vertraulicher oder proprietärer Modellarchitekturen, Parameter und Daten vor unbefugtem Zugriff Dritter.

### 4.3.2 Evaluation der Verlässlichkeit

Zur Auswertung der Ergebnisse, über die bereits vorgestellten Evaluationsmetriken hinaus, werden verlässlichkeitsbezogene Untersuchungen der Ergebnisse durchgeführt. Die Nutzung von Activation Heatmaps oder ähnlicher Methoden, die eine Visualisierung der gelernten Merkmale ermöglichen, ist eine gängige Herangehensweise, um die Transparenz von Prädiktionsergebnissen zu erhöhen [MSM18]. In diesem Kontext werden deshalb gängige Activation Heatmaps für das Modell PCB, sowie die Knoten des Graphmodells ausgegeben, um die Funktionsweise dieser Modelle transparenter zu gestalten. Eine weitere Möglichkeit zur Visualisierung stellt das Ranking der Prädiktionen dar. Auf diese Weise entwickeln Nutzer ein Verständnis dafür, weshalb falsche Zuordnungen stattgefunden haben. Diese Visualisierung schließt sich an die Metrik der Rank-k-Accuracy an. Zusätzlich werden die ausgegebenen Ergebnisse inklusive ihrer Standardabweichungen angegeben. Ohne die Angabe von Ungewissheiten oder Konfidenzintervallen, wird einem Nutzer eine irreführende Zuversicht hinsichtlich der Performance eines Modells suggeriert, welche es zu vermeiden gilt.

Zur Evaluation der erstellten Definitionen und darauf basierenden Metriken wird eine Reliabilitätsuntersuchung durchgeführt. So führen 10 Anwender und Forscher aus dem Bereich des Maschinellen Lernens die Anwendung des entwickelten Bewertungsverfahrens auf eine Auswahl von vier Modellen und vier Datensätzen durch, unter Verwendung und Kenntnisnahme der konstatierten Definitionen. Es werden hierfür ein Modell und ein Datensatz aus dieser Arbeit und drei weitreichend bekannte Modelle und Datensätze genutzt. Den Nutzern werden im Vorfeld mittels eines einleitenden Texts grundlegende Informationen zu den Datensätzen und Modellen mitgeteilt. Diese Informationstexte sowie die Auswertung der Nutzer sind im Anhang zu finden. Es handelt sich bei der Befragung um folgende Modelle und Datensätze:

- LARa (Logistic Activity Recognition Challenge) [Nie+20]: Ein Motion-Capture- und IMU-Datensatz menschlicher Aktivitäten, die von 16 Personen in einem Lagerarbeitsszenario durchgeführt wurden. Der Datensatz enthält Aktivitätsklassen typischer Lageraktivitäten, wie die Handhabung von Waren. Darüber hinaus wurden Attribute wie der Gang, die Händigkeit (links oder rechts) oder spezifische Objektpositionen annotiert. Die Annotationen bieten semantische Beschreibungen von Aktivitäten zur Unterstützung des Transferlernens. Das Aufnahmeprotokoll ist online verfügbar und die Veröffentlichung folgt dem FAIR-Prinzip [Wil+16]. Die Identitäten aller Personen wurden pseudonymisiert und anonymisiert. Alle Personen haben vor der Aufzeichnung eine Einverständniserklärung unterzeichnet.
- CelebA (Celebrity Attribute Dataset) [Liu+15]: Ein Bilddatensatz, der sich mit den Gesichtsattributen von Prominenten befasst. Die Bilder enthalten Annotationen für verschiedene Attribute wie die Haarfarbe, das Alter, das Geschlecht und Gesichtsausdrücke. CelebA wird häufig für Aufgaben wie Gesichtserkennung, Attributvorhersage und Manipulation von Gesichtsattributen in der Computer Vision Forschung verwendet. Es dient als Ressource für das Training und die Bewertung von Maschinellen Lernmodellen für gesichtsbezogene Aufgaben und verwendet öffentlich verfügbare Bilder von Personen, die weltweit bekannt sind.
- DukeMTMC (Duke University Multi-Target Multi-Camera) [Ris+16]: DukeMTMC war ein weit verbreiteter Benchmarkdatensatz für Herausforderungen beim Multi-Target Multi-Camera-Tracking (MTMC), wie die Identifikation von Personen im öffentlichen Raum. Der Datensatz wurde 2014 auf dem Campus der Duke University (USA) aufgezeichnet. Er war ein herausragender Datensatz in Bezug auf Größe, Anzahl der verwendeten Kameras und Datenvielfalt. Jedoch wurden der Datensatz und die dazugehörige Veröffentlichung seitdem zurückgezogen [PMN21], da die Aufnahmen ohne die Kenntnis der aufgezeichneten Personen erstellt und veröffentlicht wurden. Der Datensatz enthält annotierte Bounding Boxes für die aufgezeichneten Personen sowie ihre entsprechenden Identitäten, unter Verwendung verschiedener Kameraperspektiven.
- GPT-3 (Generative Pre-trained Transformer) [Bro+20]: Ein Large Language Modell (LLM), das zur Familie der Transformer Modelle gehört. GPT-3 ist auf einer großen Anzahl von Online-Textdaten trainiert worden und ist in der Lage, menschenähnlichen Text als Antwort auf Anfragen zu generieren. Das Modell ist in der Lage eine Vielzahl von Sprachaufgaben auszuführen, einschließlich Textvervollständigung, Übersetzung, Fragebeantwortung und Text- oder Codegenerierung in diversen Programmiersprachen. Weder das Modell selbst noch die darauf trainierten Daten sind Open Source verfügbar. Das Modell erlangte weit über die For-

schungsgemeinschaft hinaus große Bekanntheit und hat den Diskurs über vertrauenswürdigen Maschinelles Lernen weiter angeregt [Rut+24a].

- YOLO (You Only Look Once) [Red+16]: Ein Modell zur Detektion von Objekten, das darauf ausgelegt ist, Objekte in Bildern oder Videoframes zu erkennen, indem es das Bild in ein Raster unterteilt und Bounding Boxes sowie im Falle mehrerer Modelle Klassenwahrscheinlichkeiten für jede Rasterzelle vorhersagt. Im Gegensatz zu seinen Vorgängern erfordert YOLO keine mehrfachen Durchläufe durch das Modell und verarbeitet stattdessen das gesamte Bild in einem einzigen Forward Pass, was es schneller und effizienter macht. Aufgrund dessen wird YOLO weiterhin für diverse Aufgaben verwendet und wurde seit seiner Einführung auf diverse Weisen angepasst und erweitert.
- GoogleNet [Sze+15]: Ein CNN Modell, das von Google entwickelt wurde und für verschiedene Bildverarbeitungsaufgaben eingesetzt wird. Es wurde erstmals im Rahmen der ImageNet Large Scale Visual Recognition Challenge im Jahr 2014 vorgestellt und erzielte dabei bemerkenswerte Performance. GoogleNet führte das Konzept des Inception-Moduls ein, das mehrere Convolutions unterschiedlicher Dimensionen und Pooling-Operationen innerhalb einer Schicht integriert. Diese Struktur ermöglicht eine effiziente Erfassung sowohl lokaler als auch globaler Features in Bildern, was zu einer verbesserten Leistung bei der Klassifizierung führt.



## Kapitel 5

# Beschreibung der Datensätze und Experimente

**Zusammenfassung** In diesem Kapitel werden die erstellten Datensätze und die dazugehörigen Aufnahmesetups beschrieben. Weiter werden die beiden Identifikationsszenarien dieser Arbeit präsentiert. Anschließend wird der Begriff und das Konzept der Verlässlichkeit im Maschinellen Lernen, basierend auf den Aspekten, die im vorangegangenen Kapitel beschrieben wurden, definiert. Zuletzt wird das Vorgehen zur Quantifikation der Verlässlichkeit vorgestellt.

„To invent you have to experiment, and if you know in advance that it’s going to work, it’s not an experiment.“

– Jeff Bezos, jährlicher Amazon Aktionärsbrief (2016)

## 5.1 Erstellte Datensätze

Für die Experimente dieser Arbeit wurden drei verschiedene Datensätze erstellt. Diese wurden basierend auf den im vorangehenden Kapitel genannten Anforderungen und Ansprüchen erstellt und werden in den kommenden Abschnitten detailliert beschrieben.

### 5.1.1 Datensatz A – pallet-block-502

Datensatz A umfasst 5.020 RGB Bilder von 520 EPAL-Palettenklötzen. Die Klötze wurden aus fünf verschiedenen Perspektiven (zentriert, Rotation links und Rotation rechts, Translation links und Translation rechts) und mit zwei verschiedenen Kameras (Jai Go 5000C PGE mit Kowa LM25HC Objektiv und AV ALVIUM 1800U-1236C-CH-C mit Kowa LM35FC Objektiv) aufgenommen. Die Bildaufnahmen wurden in den Räumlichkeiten des Fraunhofer-Instituts für Materialfluss und Logistik manuell durchgeführt. Basierend auf den fünf Perspektiven und zwei Kameras sind somit 10 Bilder pro Klotz und damit pro ID vorhanden. Diese Aufnahmemodalitäten dienen der Abdeckung diverser Perspektiven und Lichtzustände, um eine Abhängigkeit dieser Umgebungsfaktoren im Rahmen des Trainings von Identifikationsmodellen zu vermeiden.

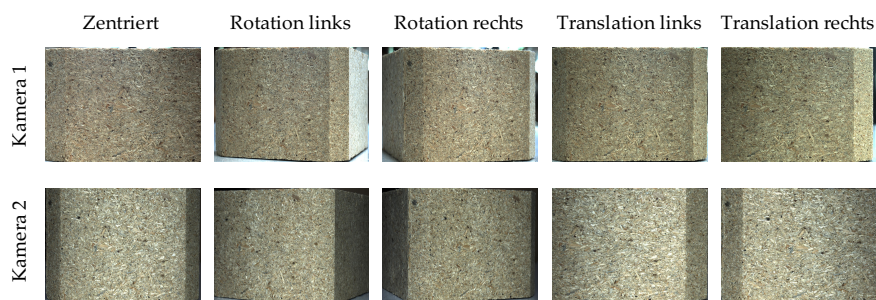


Abb. 5.1 Auszug des Datensatzes *pallet-block-502*, hier Datensatz A genannt.

Die für diesen Datensatz verwendeten Klötze sind nicht markiert, sprich sie sind nicht mit dem EPAL-Logo oder sonstigen Texten gekennzeichnet. Dies wurde ermöglicht, da die Klötze für das Forschungsprojekt „Silicon Economy Logistics Ecosystem“ von der EPAL e. V. bereitgestellt worden sind.

Dank der Exklusion industriüblicher Schriftzüge und sonstiger Markierungen, wird abermals die Abhängigkeit eines Identifikationsmodells von diesen Merkmalen vermieden. So wird im Umkehrschluss die erfolgreiche Identifikation basierend auf den einzigartigen visuellen Oberflächenmerkmalen der Klötze sichergestellt.

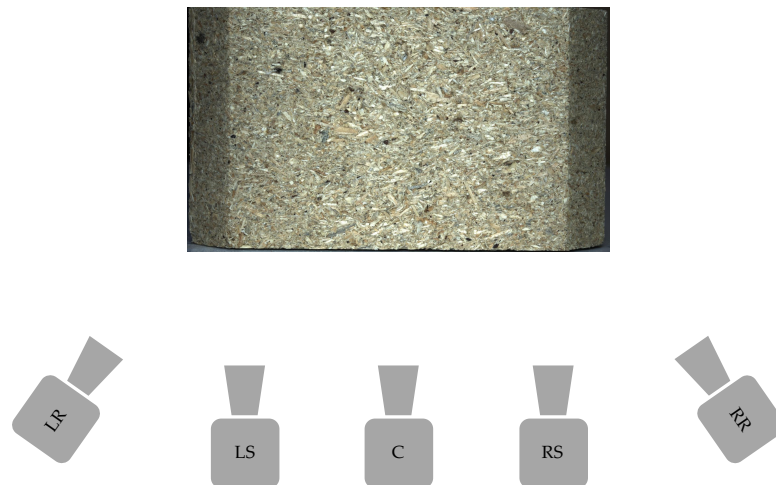


Abb. 5.2 Vereinfachte Visualisierung des Aufnahmesetups für Datensatz A.

Beispiele der bei der Erstellung von Datensatz A aufgezeichneten Bilder sind Abb. 5.1 zu entnehmen. Eine graphische Repräsentation des verwendeten Aufnahmesetups ist Abb. 5.2 zu entnehmen.

### 5.1.2 Datensatz B – pallet-block-32965

Datensatz B enthält abermals Bilder von Palettenklötzen. Diese wurden im Falle dieses Datensatzes jedoch in einem Lager eines Deutschen Großunternehmens aufgezeichnet.

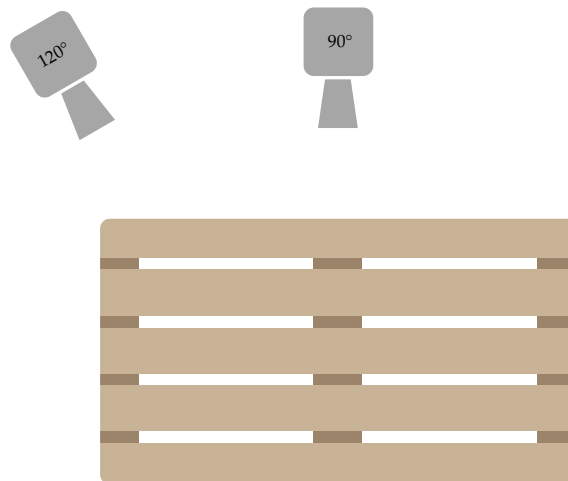
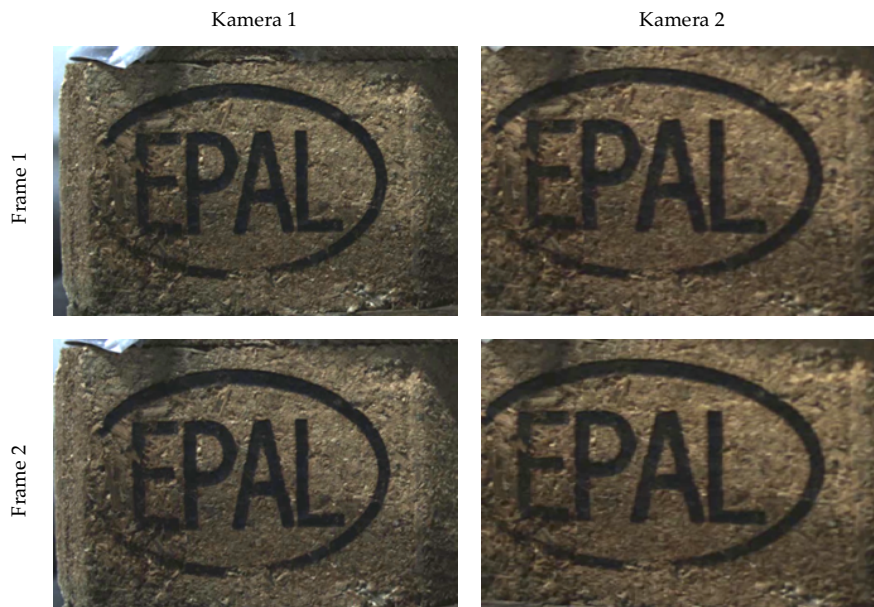


Abb. 5.3 Vereinfachte Visualisierung des Aufnahmesetups für Datensatz B.

Es wurden 131.860 Bilder von 32.965 Klötzen erstellt. Hierfür wurden dieselben Kameras wie für die Erstellung von Datensatz A genutzt. Die beiden Kameras wurden in circa 1 m Distanz zueinander und circa 0,5 m Distanz zum Förderband auf Item-Profilen aufgestellt und via Ethernet mit einem Server verbunden. Eine Kamera (Kamera 1) wurde dabei orthogonal zum Förderband aufgestellt ( $90^\circ$ ), während die andere (Kamera 2) angewinkelt zum Förderband stand ( $120^\circ$ ).



**Abb. 5.4** Auszug des Datensatzes *pallet-block-32965*, hier Datensatz B genannt.

Die Lagerumgebung verfügte während der Aufnahmen nicht über natürliches Licht. Über Kamera 2 befand sich stattdessen eine Halogenlampe, die zu einer unterschiedlichen Beleuchtung zwischen Kamera 1 und Kamera 2 führte. Die Aufnahmen wurden über einen Zeitraum von mehreren Monaten automatisiert durchgeführt, indem ein zu diesem Zeitpunkt bereits trainiertes YOLOv4 Modell die Klötze im Videostream detektierte, die relevanten Bilder entsprechend speicherte und mit einem Label versah. Die Komputation fand auf dem erwähnten Server statt, auf welchem die Bilder ebenfalls vorübergehend gespeichert wurden. Bei einer erfolgreichen Detektion wurde ein weiteres Bild mit einem Versatz von zwei Frames aufgezeichnet, weswegen vier Bilder pro ID vorhanden sind. Einige Kameraparameter (Belichtung (13.800 – 11.400 s), Gain (300 – 1.000 ISO) und Level Control (12 – 33)) wurden außerdem alle 300 s variiert. Mit diesem Aufnahmesetup und seiner Variation wird darauf abgezielt, eine Unabhängigkeit von Umgebungsbedingungen im Rahmen der Identifikation sicherzustellen. Da es sich bei dem Aufkommen

an Paletten um die tatsächlich genutzten Paletten des Unternehmens handelte, wurden nicht ausschließlich Euro-Paletten aufgezeichnet, sondern auch weitere hölzerne Paletten. Beispiele dafür sind Industrie- und UIC-Paletten. Der Aufbau des Versuchsstands ist Abb. 5.3 zu entnehmen. Beispielhafte Bilder des Datensatzes sind Abb. 5.4 zu entnehmen.

### 5.1.3 Datensatz C – galvanized-636

Datensatz C umfasst 5.088 RGB Bilder von 636 galvanisierten Stahlblechen. Die Stahlbleche wurden aus zwei verschiedenen Perspektiven ( $75^\circ$  und  $90^\circ$ ) und mit zwei verschiedenen Beleuchtungen (Umgebungslicht und Fotolicht) aufgenommen. Es sind demnach vier Bilder pro ID und damit pro Seite des Stahlblechs vorhanden. Als Kamera wurde für die Erstellung dieses Datensatzes eine Canon EOS 6D verwendet. Die Aufnahmen wurden in den Räumlichkeiten des Lehrstuhls für Förder- und Lagerwesens durchgeführt.

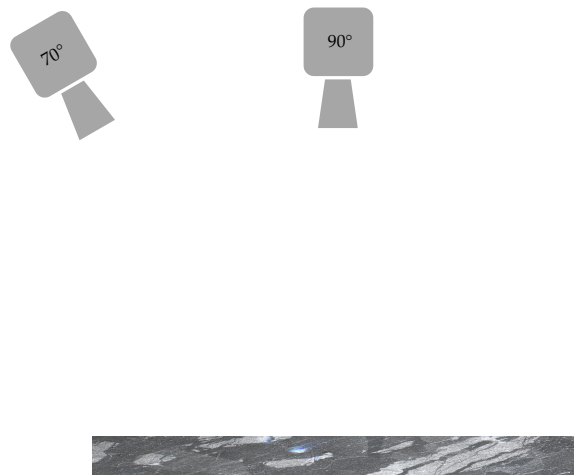


Abb. 5.5 Vereinfachte Visualisierung des Aufnahmesetups für Datensatz C.

Die Stahlbleche wurden von einer regionalen Verzinkerei eigens für die Erstellung dieses Datensatzes hergestellt und galvanisiert und sind repräsentativ für jegliche Objekte aus galvanisiertem Stahl, die sich in der industriellen Nutzung wiederfinden. Sie stehen somit auch repräsentativ für Transporthilfsmittel wie Stapeltransportkisten, deren einzigartige Oberflächenstruktur für die Identifikation verwendbar ist. Die Maße der verwendeten Stahlbleche sind  $100 \times 100 \times 3$  mm.

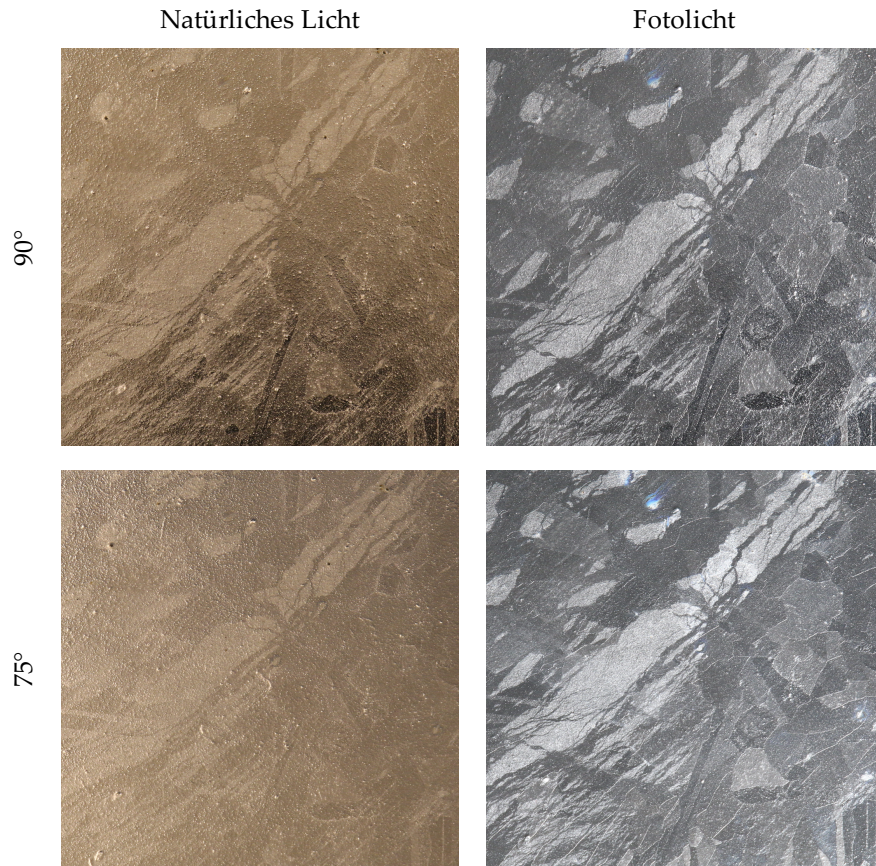


Abb. 5.6 Auszug des Datensatzes *galvanized-636*, hier Datensatz C genannt.

Beispiele der bei der Erstellung von Datensatz A aufgezeichneten Bilder sind Abb. 5.6 zu entnehmen. Eine graphische Repräsentation des verwendeten Aufnahmesetups ist Abb. 5.5 zu entnehmen.

## 5.2 Identifikationsexperimente

In diesem Abschnitt wird der Identifikationsprozess und der dafür nötige Aufbau beschrieben. Der Prozess wird in zwei Szenarien unterteilt: Szenario I – Offline Evaluation und Szenario II – Online Evaluation. Unter Nutzung dieser übergreifenden Szenarien wird die Evaluation der Identifikation unter verschiedenen Gesichtspunkten durchgeführt. Szenario I fokussiert dabei die Untersuchung der Unterschiede diverser Modelle und Datensätze und dient als Grundlage für die anschließenden Verlässlichkeitsuntersuchungen. Szenario II hingegen fokussiert die Anwendbarkeit und Nutzbarkeit in einem industrienahen Umfeld mit limitierten Hardwareressourcen und in Echtzeit.

### 5.2.1 Szenario I – Offline Evaluation (Datensatzversuch)

In Szenario I werden die ausgewählten und entwickelten Modelle auf die erstellten Datensätze angewendet. Somit werden die Datensätze zur erstmaligen Evaluation der Modelle genutzt. Die Modelle werden auf den jeweiligen Datensätzen trainiert und anschließend mittels eines Hold-out Sets evaluiert. Beide Datensätze sind zu jeweils 50 % in Gallery und Query Daten aufgeteilt. Jede ID und alle ihre zugehörigen Bilder sind nur im Trainings Datensatz oder im Hold-out Set vorhanden. Der verwendete Workflow für dieses Szenario ist Abb. 5.7 zu entnehmen.

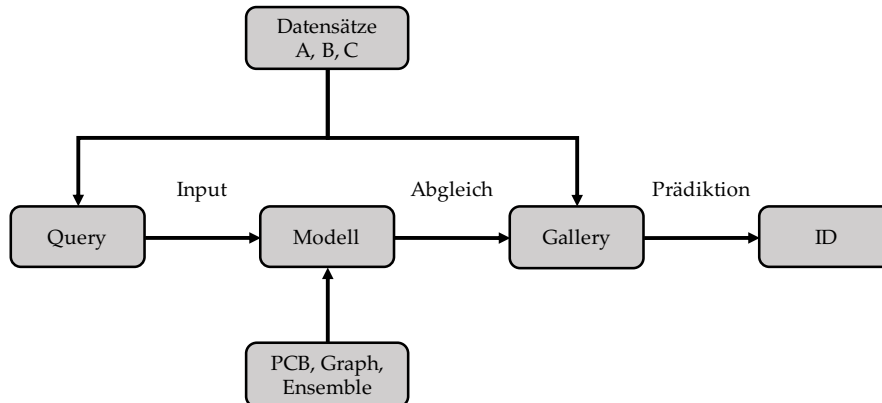


Abb. 5.7 Workflow der Identifikation für Szenario I.

Spezifisch werden also das PCB Modell, ein Graphmodell und ein Ensemblemodell verwendet. Die Trainingsparameter der jeweiligen Modelle sind Kapitel 4 zu entnehmen. Das Ensemblemodell wird mit der Konkatenation als Ensemblingmethode verwendet. Die Performance der Modelle wird anhand der Ranked Accuracy ihrer Prädiktionen evaluiert. Es wird eine fünffache Kreuzvalidierung verwendet, um Standardabweichungen der Ergebnisse anzugeben.

### 5.2.2 Szenario II – Online Evaluation (Feldversuch)

In Szenario II wird das Identifikationsverfahren in einer industrieähnlichen Umgebung (Forschungshalle des Lehrstuhls für Förder- und Lagerwesens) mobil auf einer Drohne eingesetzt und auf einen Satz von 20 Paletten angewandt (siehe Abb. 5.8). Die Paletten wurden so ausgewählt, sodass sie möglichst divers sind und die unterschiedlichen MCH-Suitability Klassen („suitable for use in machines, with conveyor technology and high-bay racking“) [EPAL 11/2021] der EPAL widerspiegeln. Somit spiegeln sie durch die Nutzung von Paletten, die diverse Abnutzungsgrade aufweisen möglichst realistisch den Stand der Nutzung in der Industrie wider. Es wurde zusätzlich

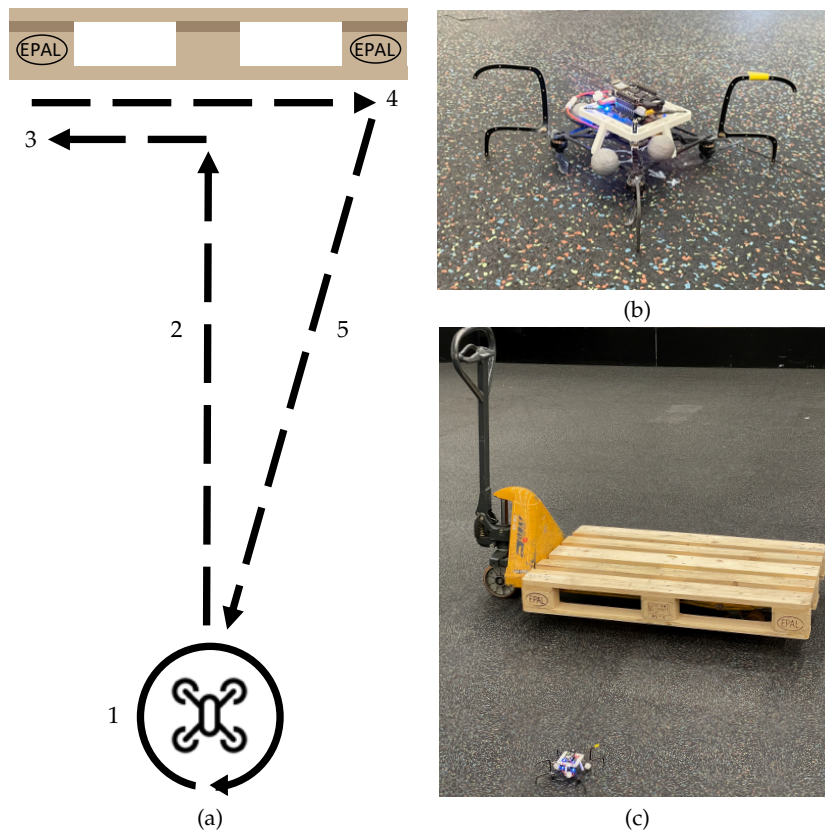
eine Industriepalette als Sonderfall verwendet, die sich von den restlichen Paletten besonders in ihren Abmessungen unterscheidet. So sind die Stirnflächen der Klötze dieser Palette quadratisch und stellen somit eine besondere Herausforderung für die autonome Navigation dar. Die Paletten werden manuell und seitenweise von ID 1 bis ID 40 annotiert.



Abb. 5.8 Übersicht der für Szenario II verwendeten Paletten.

In diesem Szenario wird eine am Standort selbst entwickelte Drohne, basierend auf der Bitcraze Crazyflie (siehe Abb. 5.9) verwendet. Die Drohne wird

mittels einer Kombination von YOLO Objekt Detektion, eines PID-Reglers und eines Vicon Motion Capture Systems und der Software CrazySwarm gesteuert. Die Drohne ist mit einer GAP8 basierten Komputationsplatine des Herstellers Bitcraze (AI Deck) ausgestattet. Das AI Deck ermöglicht die Nutzung einer on-board Kamera. Verwendet wird dabei eine Himax HM01B0-ANA RGB-Kamera mit  $324 \times 244$  px Auflösung und bis zu 60 FPS Bildwiederholungsrate. Die Drohne ist mit retroreflektiven Markern ausgestattet, um von den diversen Infrarotkameras (Vicon Vantage und Vero) des Motion Capture Systems erfasst zu werden. Das aus 52 Kameras bestehende Motion Capture System ermöglicht die Lokalisierung der Pose eines mit Markern versehenen Objekts mit einer Genauigkeit von circa 0,3 mm und überträgt diese Pose mit einer Frequenz von 300 Hz an den Server.



**Abb. 5.9** Anwendungsfall für Szenario II, dargestellt als Mission (a) und verwendete Drohne (b) in der Hallenumgebung (c).

Zur autonomen Umsetzung des Verfahrens wird der Workflow aus Abb. 5.10 verwendet. Dieser ist nötig, um die vorgeschaltete Detektion der

Paletten und ihrer Klötze zu ermöglichen. Für die Identifikation wird das im Vorfeld auf Datensatz B trainierte PCB Modell verwendet.

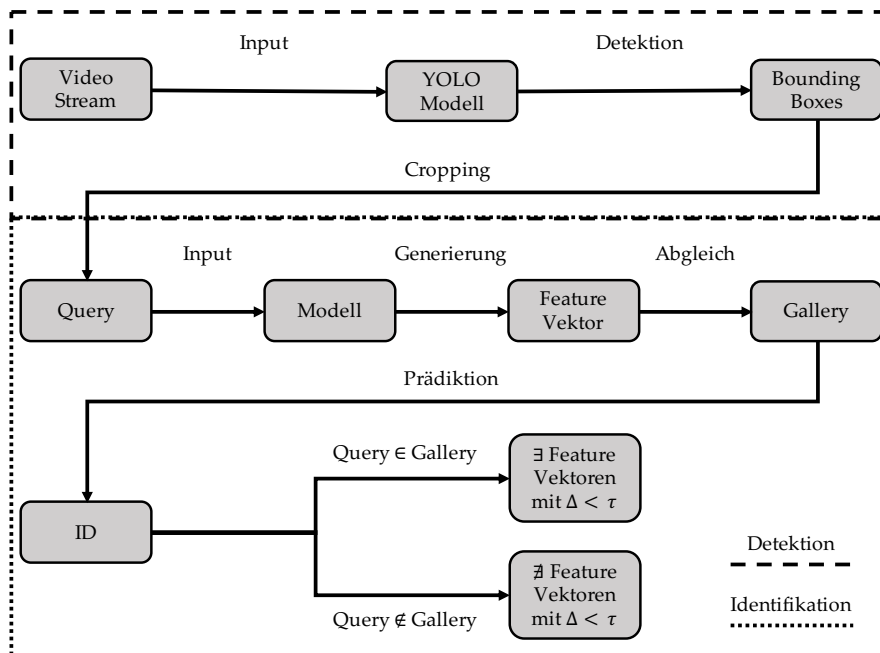


Abb. 5.10 Workflow der Identifikation für Szenario II.

Als Anwendungsfall wird ein Pufferlager simuliert. Es wird jeweils eine der Paletten auf einem Handgabelhubwagen platziert, detektiert und identifiziert. Dabei wird jeweils eine Seite der Palette (drei Klötze) verwendet. Es werden folgende Schritte (visualisiert in Abb. 5.9) durchgeführt:

1. Abheben und Rotation am Nullpunkt um bis zu  $360^\circ$ , bis eine Palette detektiert wird (YOLO Modell 1: Palettendetektion)
2. Hinflug auf den Mittelpunkt der detektierten Palette (YOLO Modell 2: Klotzdetektion) und Bilderstellung des mittleren Klotzes
3. Anflug auf den linken Klotz und Bilderstellung (YOLO Modell 2)
4. Anflug auf den rechten Klotz und Bilderstellung (YOLO Modell 2)
5. Rückflug zum Nullpunkt
6. Landung am Nullpunkt

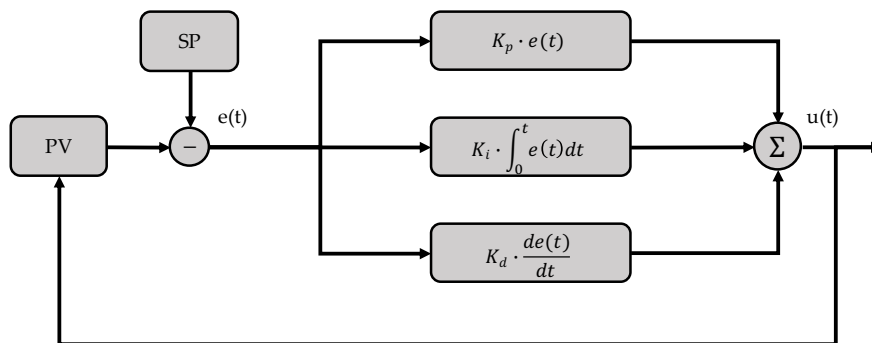
Die Modelle 1 und 2 sind jeweils zweimal trainiert worden, da sie jeweils auf Paletten angewendet werden, die nicht im Trainingsdatensatz enthalten sind (hold-out Daten). Modell 1a und 2a umfassen somit die IDs 1 - 20 und Modell 1b und 2b umfassen die IDs 21 - 40. Eine schematische und bildliche Darstellung des Anwendungsfalls und der Mission sind Abb. 5.9 zu

entnehmen. Die jeweiligen YOLOv4-Tiny Instanzen werden mit folgendem Trainings- und Testsplit und einer 80/20 Aufteilung der Daten erstellt:

**Tabelle 5.1** Datensatzaufteilung YOLOv4-Tiny (Modell 1 – Palette, Modell 2 – Klotz, a – IDs 1 - 20, b – IDs 21 - 40).

Modellart	IDs	Gesamtanzahl Bilder	Trainingsdatensatz	Testdatensatz
Palette	1 - 20	636	508	128
	21 - 40	682	554	128
Klotz	1 - 20	1.966	1.572	394
	21 - 40	1.766	1.412	354

Das Training verläuft jeweils über 1.000 Epochen unter Nutzung der cIOU Loss Funktion. Es werden die performantesten Gewichtungen des Trainings gespeichert und verwendet. Die Annotation der Bilder findet manuell, unter Nutzung der Website „makesense.ai“ statt. Die Bilder werden mit der bereits genannten Himax HM01B0-ANA RGB-Kamera der Drohne unter diversen Distanzen aufgenommen, die für den Anwendungsfall relevant sind. Die Lichtverhältnisse werden mit dimmbaren Deckenleuchten variiert und mit einem UNI-T UT383 Luxmeter zwischen 305 lx und 415 lx gemessen.



**Abb. 5.11** PID-Regelkreis der Drohne für Szenario II.

Die Steuerung der Drohne erfolgt unter Nutzung eines PID-Reglers (siehe Abb. 5.11), der die aktuelle Größe und Position der Bounding Box im Bild der Kamera der Drohne als Prozessvariable ( $PV$ ) und die gewünschte Größe und Position der Bounding Box als Führungsgröße (Setpoint,  $SP$ ) nutzt. Als Regelgröße ( $u(t)$ ) wird die relative Positionsänderung der Drohne ausgegeben. Hierbei gibt die Regelgröße die relative Positionsänderung in X-, Y- und Z-Koordinaten an.  $SP$  gibt wiederum die gewünschten Koordinaten sowie die gewünschte Größe der Bounding Box an. Der Nullpunkt des Koordinatensystems liegt hierbei in der Bildmitte und ist in X- und Y-Richtung der

gewünschte  $SP$  (die Drohne positioniert sich mittig gegenüber dem Klotz, beziehungsweise der Palette). In  $Z$ -Richtung hingegen gilt es eine Distanz zu ermöglichen, bei der die Quadratpixelfläche der Objektgröße ein Viertel der Bildfläche ergibt, da dies eine geeignete räumliche Distanz für die Identifikation darstellt. Die dafür nötige räumliche Distanz  $D$  ergibt sich approximativ über die räumliche Größe der Klötze  $o_r$ , ihrer Größe auf dem Bildsensor  $o_s$  sowie der Brennweite  $f$  der verwendeten Kamera:

$$D = \frac{o_r \cdot f}{o_s} \quad (5.1)$$

wobei sich die Brennweite bestimmen lässt als

$$f = \frac{s}{2 \cdot \tan\left(\frac{\theta}{2}\right)} \quad (5.2)$$

mit dem Kamerawinkel  $\theta$  und der Sensorgröße  $s$ .

Die Größe der Klötze auf dem Bildsensor  $o_s$  lässt sich wiederum auf Basis der gewünschten Größe des Klotzes  $o_g$  auf der Bildfläche bestimmen. Ziel ist es hier, die Bildfläche zu einem Viertel abzudecken, also eine Quadratpixelfläche von  $162 \times 122$  px, bei einer Bildabmessung  $b$  von  $324 \times 244$  px. So lässt sich  $o_s$  bestimmen als:

$$o_s = o_g \cdot \frac{s}{b} \quad (5.3)$$

wobei das Verhältnis von  $s$  zu  $b$  hier die Pixelgröße  $g$  bestimmt.

Die Abmessungen eines EPAL-Klotzes sind  $145 \times 100$  mm, der Kamerawinkel beträgt  $87^\circ$  und die Sensorgröße liegt bei  $1,17 \times 0,80$  mm. So ergibt sich folgende Zieldistanz:

$$D = \frac{o_r \cdot \frac{s}{2 \cdot \tan\left(\frac{\theta}{2}\right)}}{o_g \cdot \left(\frac{s}{b}\right)} = \frac{145 \text{ mm} \cdot \frac{1,17 \text{ mm}}{2 \cdot \tan\left(\frac{87^\circ}{2}\right)}}{162 \text{ px} \cdot \frac{1,17 \text{ mm}}{324 \text{ px}}} = 152,2 \text{ mm} \quad (5.4)$$

Die Linsendistortion wird bei der Berechnung vernachlässigt.

Wie in Kapitel 2 beschrieben, bietet sich der PID-Regler für die Nutzung in diesem Kontext besonders an. Neben dem P-Anteil des Reglers eignet sich der I-Anteil für eine zunehmende Korrektur der Flugposition über den Verlauf der Zeit und der D-Anteil während Bewegungsabschnitten mit hohen Beschleunigungsraten. So wird für die Mission eine kontrollierte Regelung und Steuerung umgesetzt. Algorithmisch ist der Regelkreis ergänzend wie folgt dargestellt:

**Algorithmus 3** Drohnenregelung mittels PID-Regler.

---

**Input:** Setpoint (Soll-Wert)  $SP$ , Prozessvariable (Ist-Wert)  $PV$   
**Output:** Kontrollvariable  $u(t)$

Define  $K_p, K_i, K_d$  ▷ Reglerparameter  
Initialize  $i(0) \leftarrow 0$

**while true do**

$e(t) \leftarrow SP - PV$  ▷ Berechnungsfehler  
 $p(t) \leftarrow K_p \cdot e(t)$  ▷ P-Anteil  
 $i(t) \leftarrow K_i \cdot \int_0^t e(t) dt$  ▷ I-Anteil  
 $d(t) \leftarrow K_d \cdot \frac{de(t)}{dt}$  ▷ D-Anteil  
 $u(t) \leftarrow p(t) + i(t) + d(t)$  ▷ Output bestimmen  
Apply  $u(t)$  to the process ▷ Steuerung umsetzen  
 $PV \leftarrow$  Read new process variable  
 $t \leftarrow t + dt$  ▷ Update Zeitschritt

**end while**

---

Um repräsentative, signifikante Ergebnisse zu erhalten, gilt es eine Anzahl Versuche durchzuführen, die ein Konfidenzniveau von 95 % (Signifikanz  $\alpha = 0,05$ ) ermöglicht. Die höchste Varianz  $\sigma$  wird bei einer empirischen Identifikationsgenauigkeit von 50 % angenommen (sprich, jede zweite Entität würde korrekt identifiziert werden), weswegen  $\sigma = 0,5$  gewählt wird ( $p \cdot q$  mit  $q = p - 1$  und Toleranz  $d = 0,05$ ). Da bei dem Versuch von einer Normalverteilung auszugehen ist, ergibt sich die Berechnung der Versuchsanzahl (Stichprobengröße  $n$ ) wie folgt:

$$n = qnorm(1 - \alpha)^2 \cdot \frac{\sigma^2}{d^2} = 270,5543 \quad (5.5)$$

Aus diesem Grund müsste die Mission aus Abb. 5.9 in Summe 271 Mal durchgeführt werden, was eine Aussagekraft der empirisch bestimmten Identifikationsgenauigkeit mit einer Konfidenz von 95 % ermöglicht. Um eine gleiche Anzahl an Flügen pro Palette zu ermöglichen, wird die Mission 280 Mal durchgeführt, womit die 40 Palettenseiten jeweils siebenmal angefliegen werden. Zuletzt ist eine Übersicht aller Softwarekomponenten für Szenario II aus Abb. 5.12 zu entnehmen.

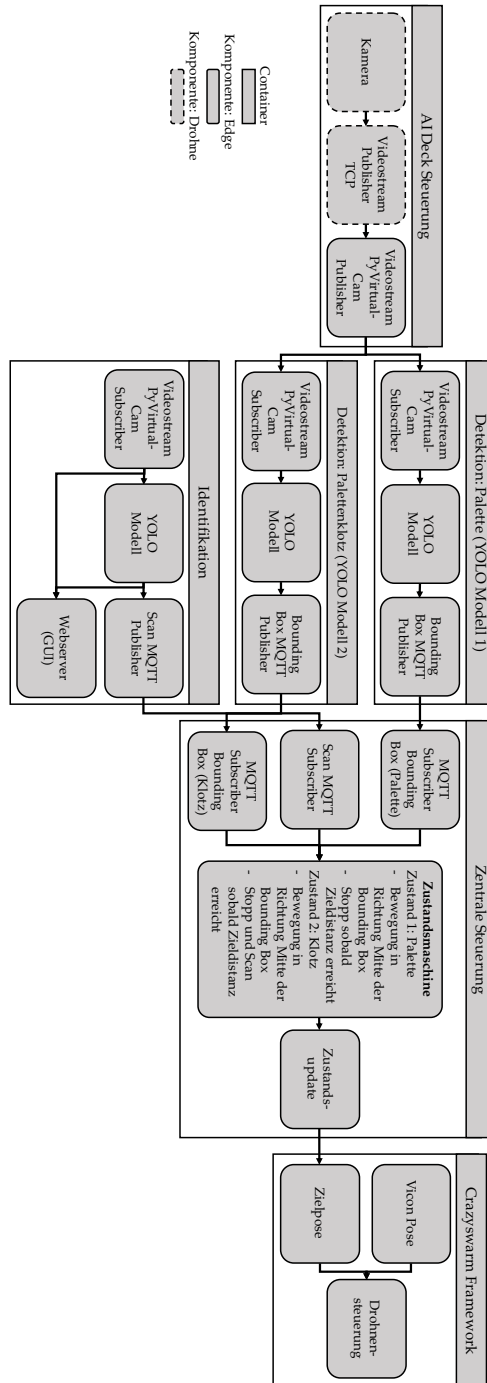


Abb. 5.12 Übersicht der Softwarekomponenten für Szenario II.

## 5.3 Verlässlichkeit der Identifikation

In diesem Kapitel werden die Ergebnisse dieser Arbeit hinsichtlich der Verlässlichkeit der Identifikation dargestellt. Dabei wird zunächst die erarbeitete Definition des Begriffs der Verlässlichkeit vorgestellt. Im Anschluss wird sich auf die Quantifikation der Verlässlichkeit fokussiert, wobei die erstellte Metrik präsentiert und verwendet wird. Darüber hinaus werden zusätzliche Aspekte der Verlässlichkeit der in den vorangehenden Kapiteln dargestellten Ergebnissen diskutiert.

### 5.3.1 Definition des Begriffs der Verlässlichkeit

Basierend auf der Recherche hinsichtlich der Verlässlichkeit im Kontext des Maschinellen Lernens wurden die Aspekte der Fairness, Robustheit, Integrität, Erklärbarkeit und Sicherheit als Schlüsselkomponenten aus der Literatur deduziert. Darauf aufbauend wurden operationelle Definitionen der jeweiligen Komponenten und folgende Konzeptionalisierung erarbeitet:

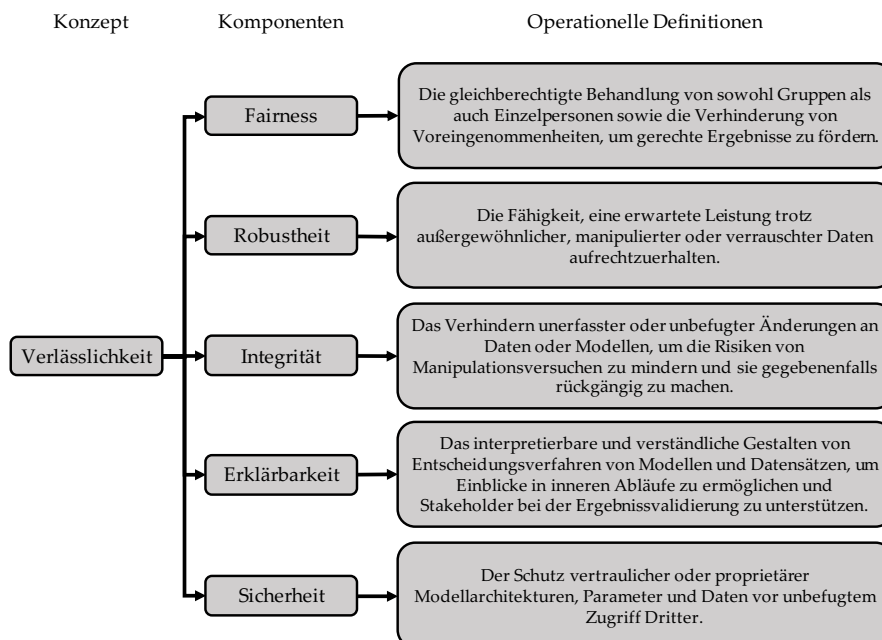


Abb. 5.13 Operationalisierung des Begriffs und Konzepts der Verlässlichkeit.

Aufbauend auf dieser Konzeptionalisierung und den zugrundeliegenden Definitionen der Schlüsselkomponenten, wird die Verlässlichkeit im Kontext des Maschinellen Lernens operationell wie folgt definiert:

**Definition:** *Verlässlichkeit*

Das Konzept der Verlässlichkeit im Maschinellen Lernen umfasst den fairen Umgang mit Daten, eine robuste Leistungserbringung beim Auftreten anomaler Daten, die Sicherstellung der Integrität von Daten und Modellen, die Bereitstellung erklärbarer Entscheidungen sowie die sichere Nutzung vertraulicher Informationen.

### 5.3.2 Quantifikation der Verlässlichkeit

Unter Anbetracht der nun vorhandenen Definitionen für sowohl die Aspekte der Verlässlichkeit, als auch des Konzepts der Verlässlichkeit selbst, ist im Folgenden eine Operationalisierung umzusetzen. Da auch im Kontext der Bewertung der Verlässlichkeit des Maschinellen Lernens die Bewertung eines Konzepts, Ergebnisses oder Prozesses stattfindet, lassen sich Analogien zu Qualitätssicherungsverfahren beobachten. Aus diesem Grund bietet sich die Adaptierung bestehender Qualitätssicherungsverfahren für die Konzeption einer Quantifikation der Verlässlichkeit an. [Sch+22] haben auf diese Weise ihr KI-Risikoschema konzeptionalisiert. Ein prominentes Modell, das im Zusammenhang mit Qualitätssicherung verwendet wird, ist die Fehlermöglichkeits- und Einflussanalyse (FMEA). FMEA ist eine systematische und halbqualitative Risikoanalyse-Methode [Sta19]. Die Verwendung von FMEA wird in vielen Standards empfohlen, wie beispielsweise im Qualitätsmanagement [Liu+23]. Je nach Anwendungsbereich wird FMEA verwendet, um Systeme, Software oder Prozesse zu bewerten [Liu+23; Sta19].

**Tabelle 5.2** Skalen der Auftretenswahrscheinlichkeit (Occurrence, O), der Signifikanz (Significance, S) und der Detektionswahrscheinlichkeit (Detection, D) eines Verlässlichkeitsrisikos eines Datensatzes oder Modells des Maschinellen Lernens.

Occurrence (O)		Significance (S)		Detection (D)	
Wahrscheinlichkeit		Auswirkung		Wahrscheinlichkeit	
Unmöglich	10	Vernachlässigbar	10	Garantiert	10
Unwahrscheinlich	9	Kaum bemerkbar	9	Hoch	9
Sehr niedrig	7 - 8	Unbedeutend	7 - 8	Moderat	7 - 8
Niedrig	4 - 6	Moderat	4 - 6	Niedrig	4 - 6
Moderat	2 - 3	Schwerwiegend	2 - 3	Sehr niedrig	2 - 3
Hoch	1	Sehr schwerwiegend	1	Unwahrscheinlich	1
Garantiert	0	Inakzeptabel	0	Unmöglich	0

Im Kern der FMEA Methode liegt die präventive Detektion möglicher Fehler und die damit verbundenen Ursachen eines anschließenden Ausfalls

eines Systems oder Vorgehens. Anwender der Methode werden bei der Nutzung aufgefordert, für jeden Ausfalltypen spezifische Fehler sowie deren jeweilige Konsequenzen und Ursachen aufzulisten. Diese Einschätzung ist subjektiver Natur und beruht auf der Qualifikation der Anwender. Für jeden Fehler werden die Auftretenswahrscheinlichkeit (Occurrence  $O$ ), die Signifikanz des Fehlers (Significance  $S$ ) und die Detektionswahrscheinlichkeit (Detection  $D$ ) auf einer Skala von 1 bis 10 bewertet (OSD-Bewertung). Höhere Zahlen repräsentieren hier eine erhöhte Auftretenswahrscheinlichkeit, eine höhere Signifikanz und eine niedrigere Detektionswahrscheinlichkeit. Die resultierenden Werte werden miteinander multipliziert und münden in einer Risikoprioritätszahl (Risk Priority Number RPN) zwischen 1 und 1.000 für jeden Fehler. Eine höhere RPN stellt ein höheres Qualitätsrisiko dar [Sta19]. Für die individuellen RPN wird eine Maßnahme zur Risikominderung vorgeschlagen und nach ihrer Implementierung wird eine neue RPN pro Fehler berechnet. Somit wird keine Gesamtwertung berechnet und der Fokus liegt auf der Empfehlung von Maßnahmen zur Qualitätsverbesserung, welche nach ihrer Implementierung evaluiert werden. Wenn die RPN gesenkt wurden, wird von einer geglückten Risikominderung ausgegangen. Wenn diese nicht gesenkt wurden, werden weitere Maßnahmen eingeleitet.

**Tabelle 5.3** Exemplarische Bestimmung des Verlässlichkeitswerts  $T$  in tabellarischer Darstellung.

Aspekt	Risiko	$O$	$S$	$D$	$\Pi$	$\bar{\Pi}$	$\omega$	$T_\omega$
Fairness	Inputs voreingenommen angefordert	4	4	8	5,04	5,04	0,2	1,01
Robustheit	Risiko von inversion attacks	4	8	9	6,6	5,89	0,2	1,18
	Risiko von adversarial attacks	7	4	5	5,19			
Integrität	Das Modell ist nicht Open Source	3	9	2	3,78	3,78	0,2	0,76
Erklärbarkeit	Illusion der Erklärungstiefe	8	4	5	5,43	5,43	0,3	1,63
Sicherheit	Unzureichende Zugriffskontrolle	6	3	6	4,76	4,76	0,1	0,48
							$T$	5,06

Für die Quantifikation des Konzepts der Verlässlichkeit eignet sich die Adaptierung eines Vorgehens angelehnt an die FMEA Methode. Dies liegt an der differenzierten OSD-Bewertung von Ausfällen, die qualitative Bewertungen (die Einschätzung der Anwender) in quantitative Werte umsetzt (die RPN und ihre konstituierenden Werte). Allerdings sind einige Anpassungen an diesem Vorgehen erforderlich. In einem ersten Schritt ist es sinnvoll, Ausfälle als Teile der zuvor definierten Aspekte der Verlässlichkeit, also der Fairness, Robustheit, Integrität, Erklärbarkeit und Sicherheit zu verstehen. Anstatt potenzieller Ausfälle werden demnach potenzielle Einschränkungen dieser definierten Aspekte betrachtet. Dies umfasst Phänomene, die den jeweiligen

Aspekt gefährden würden und die demnach als Risiken hinsichtlich eines verlässlichen Modells oder Datensatzes zu betrachten sind. Dazu wird den Nutzern eine Tabelle mit Risiken pro Verlässlichkeitsaspekt zur Verfügung gestellt. Diese sind jeweils entweder auf Datensätze oder Modelle anzuwenden. Die Nutzer haben eine Auswahl von ein bis drei Risiken pro Aspekt, auf deren Basis der jeweils resultierende Wert pro Aspekt berechnet wird. Die aufgelisteten Risiken wurden aus der Literatur deduziert (siehe Kapitel 4) und sind Tab. 5.4 zu entnehmen.

---

**Algorithmus 4** FRIES Trust Score (Verlässlichkeitswerts) Algorithmus.
 

---

**Input:**

$\omega_i \forall i \in [0,5); \omega_i \geq 0.1$	▷ Gewichtung der Aspekte
$\Psi_i^j \forall i \mid 0 \leq j < n_i \mid 1 \leq n_i \leq 3$	▷ Auswahl der Risiken pro Aspekt
$O_{\Psi_i^j} \forall i, j; O_{\Psi_i^j} \in [0,10]$	▷ Auftretenswahrscheinlichkeit
$S_{\Psi_i^j} \forall i, j; S_{\Psi_i^j} \in [0,10]$	▷ Signifikanz
$D_{\Psi_i^j} \forall i, j; D_{\Psi_i^j} \in [0,10]$	▷ Detektionswahrscheinlichkeit

**Output:** Resultierender Trust Score  $0 \leq T \leq 10$  $sum_{\omega} \leftarrow \sum_i \omega_i$  $\omega_i \leftarrow \frac{\omega_i}{sum_{\omega}}$ **for each**  $i \in [0,5)$  **do**  **for each**  $j \in [0, n_i)$  **do**     $T_i^j \leftarrow \sqrt[3]{O_{\Psi_i^j} \cdot S_{\Psi_i^j} \cdot D_{\Psi_i^j}}$     **if**  $O_{\Psi_i^j} = 10 \vee S_{\Psi_i^j} = 10 \vee D_{\Psi_i^j} = 10$  **then**       $T_i^j \leftarrow 10$     **end if**    **if**  $O_{\Psi_i^j} = 0 \vee S_{\Psi_i^j} = 0 \vee D_{\Psi_i^j} = 0$  **then**       $T_i^j \leftarrow 0$     **end if**  **end for**   $T_i \leftarrow \frac{1}{n_i} \sum_{j=0}^{n_i-1} T_i^j$   **for each**  $j \in [0, n_i)$  **do**    **if**  $T_i^j = 0$  **then**       $T_i \leftarrow 0$     **end if**  **end for****end for** $T \leftarrow \sum_{i=0}^4 \omega_i \cdot T_i$ **return**  $T \in [0,10]$ ▷ Trust Score  $T$

Auf dieser Grundlage wird die qualitative Bewertung nun mittels eines angepassten OSD-Vorgehens in eine quantitative Bewertung umgewandelt. Des Weiteren wird die Skala um den Wert 0 ergänzt. Dieser ist beispielsweise nötig in Situationen, in denen die Auftretenswahrscheinlichkeit garantiert ist, die Signifikanz eines Risikos unannehmbare Auswirkungen hat oder die Detektion des Risikos unmöglich ist. Eine Übersicht der adaptierten OSD-Bewertung und damit der Skalen, die den Nutzern zur Auswahl gestellt werden, ist Tab. 5.2 zu entnehmen.

Anders als beim FMEA Verfahren werden die resultierenden OSD-Werte anschließend miteinander multipliziert, was in einem Wert  $\bar{\Pi}$  pro Aspekt resultiert. Die fünf resultierenden Werte werden standardmäßig zu gleichen Teilen à je 20 % berücksichtigt. Die Werte sind je nach Anwendungsfall von den Nutzern unterschiedlich zu gewichten, mit einer Mindestgewichtung von 10 % pro Aspekt. Basierend auf den gewichteten Werten  $T_{\omega}$  wird der finale Verlässlichkeitswert  $T$  (FRIES Trust Score) berechnet. Dabei werden Werte innerhalb eines Bewertungsschemas von 0 bis 10 berechnet, wobei höhere Werte eine höhere Verlässlichkeit darstellen. Ein hoher Wert ist somit nicht repräsentativ für hohe Risiken, sondern für ein vertrauenswürdigen Modell oder einen vertrauenswürdigen Datensatz. Da die FMEA Methode tendenziell zu eher kleinen Werten führt (mit zufälligen Werten für  $O$ ,  $S$  und  $D$ , sind 92 % der FMEA Werte  $\leq 4,5$ ), wird als Kombinationsfunktion anstatt einer gängigen Multiplikation der Werte ( $O \cdot S \cdot D$ ) die Kubikwurzel ebenjener Multiplikation verwendet:  $\sqrt[3]{O \cdot S \cdot D}$ . Diese Berechnung ermöglicht eine probabilistische Gleichverteilung der erhaltenen Werte. Bei der Auswahl eines optimalen Werts von 10 für  $O$ ,  $S$ , oder  $D$  wird der jeweilige Aspekt in Gänze auf 10 gesetzt. Dies ermöglicht es beispielsweise ein Risiko zu ignorieren, dessen Auftreten auszuschließen ist (wenn  $O = 10$ , ist selbst eine Signifikanz von  $S = 1$  irrelevant). Analog ist es möglich ein Defizit durch ein absolutes Ausschlusskriterium, mit einem Wert von 0 auszuweisen. Hier bedeutet ein Defizit bei auch nur einem der Aspekte: Das Modell oder der Datensatz sind als Ganzes nicht als verlässlich einzuschätzen. Eine Übersicht der Berechnungsmatrix für das vorgeschlagene Verfahren ist in Tab. 5.3 exemplarisch zu sehen. Des Weiteren ist eine formale Darstellung des Berechnungsansatzes Alg. 4 zu entnehmen.

Tabelle 5.4 Verlässlichkeitsrisiken für Modelle und Datensätze des Maschinellen Lernens.

Aspekt	Risiko
Fairness	Entscheidungen, die vom Modell getroffen werden, sind gegen bestimmte Gruppen oder Einzelpersonen voreingenommen
	Benutzereingaben werden auf voreingenommene Weise angefordert
	Die Leistung weicht für gewisse Gruppen ab oder ist nur auf bestimmte Gruppen anwendbar
	Der Datensatz repräsentiert nicht die Anwendung (Sampling Bias)
	Der Datensatz enthält geschützte Attribute
	Der Datensatz verbreitet Vorurteile (da er etwa aus ungefilterten Webdaten generiert wird)
Erklärbarkeit	Der Entscheidungsprozess des Modells ist intransparent
	Die Architektur des Modells ist unbekannt oder hindert ihre Interpretierbarkeit
	Die Ausgaben des Modells sind nicht durch Stakeholder validierbar
	Fehlende Dokumentation des Datenakquise- und Annotationsprozesses
	Der Datensatz ist für Menschen nicht verständlich
	Es besteht Unklarheit darüber, wie fehlende Werte oder Ausreißer im Datensatz behandelt werden
Sicherheit	Entscheidungen oder interne Repräsentationen würden sensible Informationen offenlegen
	Unzureichende Zugriffskontrolle zu proprietärem Modell
	Fehlerhafte Entscheidungen würden zu kritischen Konsequenzen führen
	Unzureichende Zugriffskontrolle zu proprietären Daten
	Offenlegung sensibler Informationen durch Metadaten oder Auxiliardaten
	Mangel an transparenten Richtlinien für Datenverwaltung (z. B. Datenverwendungsvereinbarungen)
Robustheit	Risiko von Adversarial oder Inversion Attacks nicht vorgebeugt
	Das Modell generalisiert nicht auf unterschiedliche Datensätze
	Wiederholte Ausführungen des Modells erzeugen nicht dieselben oder ähnliche Ausgaben
	Der Datensatz enthält keine Grenzfälle oder Ausreißer
	Die Daten sind anfällig für Distribution Shifts
	Die Daten enthalten schädliche Anomalien oder Perturbationen
Integrität	Es kann nicht garantiert werden, das Modell sei nicht manipuliert worden
	Es werden keine Ausgabeunsicherheiten angegeben
	Änderungen am Modell sind nicht nachverfolgbar
	Es kann nicht garantiert werden, die Daten seien nicht manipuliert worden
	Änderungen am Datensatz sind nicht nachverfolgbar
	Signifikante Labelingunsicherheiten sind nicht ausgeschlossen

## Kapitel 6

# Evaluation der Experimente

**Zusammenfassung** In diesem Kapitel werden die Ergebnisse dieser Arbeit evaluiert. Dabei wird auf die erstellten Datensätze sowie die Identifikationsgenauigkeit innerhalb der jeweiligen Szenarien eingegangen. Des Weiteren wird die Verlässlichkeitsquantifikation hinsichtlich ihrer Güte und Reliabilität evaluiert. Abschließend wird der Erkenntnisgewinn der Arbeit kritisch gewürdigt.

„In God we trust. All others must bring data.“  
– Mary Walton, „The Deming Management Method“, Kapitel XX (1986)

## 6.1 Datensatzqualität

In diesem Kapitel werden die in dieser Arbeit erstellten Datensätze hinsichtlich ihrer Qualität evaluiert. Es werden dabei Aspekte wie die Annotationsgenauigkeit, der Datensatzumfang und die industrielle Anwendbarkeit der teil-automatisiert aufgezeichneten Daten untersucht. Wie in Kapitel 4.1 detailliert, findet die Evaluation der Datensätze auf Basis der Anforderungen der Identifikation sowie grundlegender Datensatzqualitätsbetrachtungen statt. Dieses Vorgehen ist angelehnt an die Arbeit von Pipino et al. [PLW02].

Die Datengrundlage für das Training der YOLO Modelle wird in diesem Kapitel nicht näher untersucht, da lediglich die Qualität der Identifikationsdatensätze für die Identifikation von Relevanz ist. Die Trainingsdaten der YOLO Modelle und ihre zugehörige Dokumentation sind jedoch in den im Anhang angefügten Quellen und Repositories einsehbar. Eine implizite Evaluation der Datensatzqualität findet im Rahmen dieser Arbeit als Bestandteil der Evaluation von Szenario II statt. Die erfolgreiche Durchführung der Mission des Szenarios setzt eine gelungene Detektion von Palette und Klötzen voraus, welche wiederum einen adäquaten Trainingsdatensatz voraussetzt.

### 6.1.1 Evaluation von Datensatz A

Datensatz A stellt die von der EPAL e. V. bereitgestellten Palettenklötze in hoher Auflösung und unter ausgeleuchteten Umgebungsbedingungen dar. Es herrscht eine hohe Variation zwischen den visuellen Perspektiven und es wurden für die Aufnahmen zwei Kameras verwendet, wodurch in Summe 10 Bilder pro Klotz aufgenommen wurden. Die Aufnahmeparameter ermöglichen somit diverse Vergleiche, selbst zwischen den Bildern einer einzelnen ID. Die Bilder und die physischen Klötze selbst sind manuell annotiert worden. Die Annotation der physischen Klötze bietet eine zusätzliche Form der Vergewisserung, hinsichtlich der korrekten Durchführung des Annotationsprozesses.

Der Datensatz ist dennoch verhältnismäßig klein, da er lediglich 502 IDs umfasst. Seine Größe von 5.020 Bildern erreicht er lediglich, da 10 Bilder pro ID vorhanden sind. Zusätzlich ist die Darstellung der Klötze wenig realistisch, da in der industriellen Praxis nur für einen kurzen Teil ihres Lebenszyklus makellose Paletten verwendet werden. Tatsächlich spielen in der Praxis Faktoren wie Verstaubung, Verschmutzung und Beschädigung eine große Rolle, die das Erscheinungsbild der Klötze kurz nach dem ersten Gebrauch verändern. Darüber hinaus sind keine Bilder von Klötzen vorhanden, die mit EPAL-Schriftzügen gekennzeichnet sind. Eine solche Kennzeichnung wäre unter Verwendung von Euro-Paletten jedoch regelmäßig vorhanden. Allerdings ist zwecks Abwesenheit von Schriftzügen bei Nutzung dieses Datensatzes, der Ausschluss der Schrift als relevantes Identifikationsfeature

sichergestellt. Zuletzt lässt sich eine eingeschränkte Variation der Belichtung beobachten, die möglicherweise ein Overfitting auf die verwendeten Lichtverhältnisse verursachen.

### 6.1.2 Evaluation von Datensatz B

Datensatz B steht in direktem Vergleich mit Datensatz A, da die gleichen Identifikationssubjekte untersucht werden. Dennoch ist der Modus Operandi der Aufnahme dieses Datensatzes ein gänzlich anderer, als der von Datensatz A. Durch die monatelange, automatisierte Aufnahme der Bilder in den Lagerstätten eines Industrieunternehmens, wurden sowohl eine größere Menge an Bildern, als auch eine realistischere, praxistauglichere Darstellung der Klötze erzielt. Datensatz B umfasst somit etwa 65 Mal mehr IDs als Datensatz A (wenn auch nur etwa 26 Mal mehr Bilder, da weniger Perspektiven pro ID vorhanden sind). Die jeweiligen Paletten beschränken sich dabei nicht nur auf Euro-Paletten, sondern auch auf andere, gängige Palettenformate. Diese bestehen teilweise aus Vollholz, sind beschädigt, verschmutzt, oder repariert (durch Reparaturnägeln gekennzeichnet).

Um eine automatisierte Datensatzaufnahme zu gewährleisten, wurde das bereits vorgestellte YOLO Modell verwendet. Da die von diesem Modell erstellte Bounding Box als Prädiktion der 2D-Position des Klotzes zu verstehen ist, ist diese approximativer Natur und somit stellenweise weniger präzise, als eine manuelle Annotation. Dies bedeutet im Umkehrschluss, die Aufzeichnung der Bilder beinhaltet teilweise Hintergründe, wenn ein zu großer Teil des Bildes ausgeschnitten wurde. Auch ist womöglich ein Klotz nicht in Gänze sichtbar, wenn ein zu kleiner Teil des Bildes ausgeschnitten wurde. Diese Einschränkung hat folglich Einflüsse auf die Nutzbarkeit des Datensatzes und beeinflusst die Identifikationsergebnisse. Eine weitere Einschränkung des Datensatzes stellt das Auftreten unscharfer Bilder dar. Diese treten in wenigen Fällen auf, in welchen die Positionierung der Palette auf dem Förderband deutlich von der durchschnittlichen Positionierung abwich und somit die eingestellte Tiefenschärfe einer oder beider Kameras nicht für eine scharfe Aufnahme geeignet war. Darüber hinaus stellt die vom Detektionsmodell kalkulierte Bounding Box nur einen Teil des Bildausschnitts dar und legt die resultierende und somit von Bild zu Bild variierende Auflösung fest. Die Auflösung der Bilder des Datensatzes ist deshalb deutlich niedriger, als die Auflösung der Bilder aus Datensatz A. Des Weiteren hat die Annotation dieses Datensatzes automatisiert stattgefunden und die im Datensatz enthaltenen Paletten gehören zum Palettenkreislauf des Unternehmens, weshalb sie nicht mehr aufzufinden sind. Ein Nachjustieren der erfolgten Bildserien oder ein Abgleich der Bilder ist somit nicht möglich. Die automatisierte Annotation hat außerdem in einigen wenigen Fällen zu fehlerhaften Annotationen geführt, was im Nachgang untersucht und korrigiert wurde.

Zuletzt gilt es die im Vergleich zu Datensatz A geringere Variationen der Perspektiven zu nennen, die pro ID vorhanden sind. Die Lichtverhältnisse sind hier durch die künstliche Beleuchtung der Lagerumgebung vorgegeben gewesen und variieren leicht, jedoch nicht in einem signifikanten Ausmaß.

### 6.1.3 Evaluation von Datensatz C

Datensatz C unterscheidet sich von Datensatz A und Datensatz B grundsätzlich, da er einen anderen Werkstoff darstellt. Der Datensatz weist eine hohe Auflösung und dank unterschiedlicher Beleuchtung und Perspektiven eine mittlere Variation zwischen den Bildern je ID auf. Wie bei Datensatz A erfolgte der Annotationsprozess manuell und die IDs wurden physisch aufgetragen.

Der Datensatz weist in etwa die gleiche Größe wie Datensatz A auf, repräsentiert jedoch jede ID mit vier anstatt 10 Bildern. Somit sind trotz der ähnlichen Datensatzgröße deutlich mehr IDs in diesem Datensatz enthalten. Da der Datensatz abermals unter Laborbedingungen aufgenommen worden ist, ist auch dieser Datensatz weniger realistisch als Datensatz B. Dies trifft etwa hinsichtlich Verschmutzung, Beschädigung und Belichtung zu. Der in Datensatz C verwendete Werkstoff ist im Vergleich zu Datensatz A und Datensatz B weniger anfällig gegenüber Beschädigungen. Belichtungsunterschiede haben jedoch einen größeren Einfluss auf das Erscheinungsbild des Werkstoffs, da der galvanisierte Stahl Lichteinstrahlung reflektiert. Der Datensatz bietet sich als Grundlage für ein Proof of Concept der Identifikation von galvanisiertem Stahl an und ist als analog zu Datensatz A zu verstehen. Ähnlich wie in Datensatz B wäre eine industrienähere Aufnahme des Werkstoffs eine relevante Ergänzung.

## 6.2 Identifikationsgenauigkeit

In diesem Kapitel werden die Ergebnisse der bereits beschriebenen Test-szenarien ausgewertet. Dabei werden die Identifikationsgenauigkeiten pro Modelle und Datensätze untersucht sowie übergreifende Herausforderungen bei der Durchführung von Szenario II beschrieben.

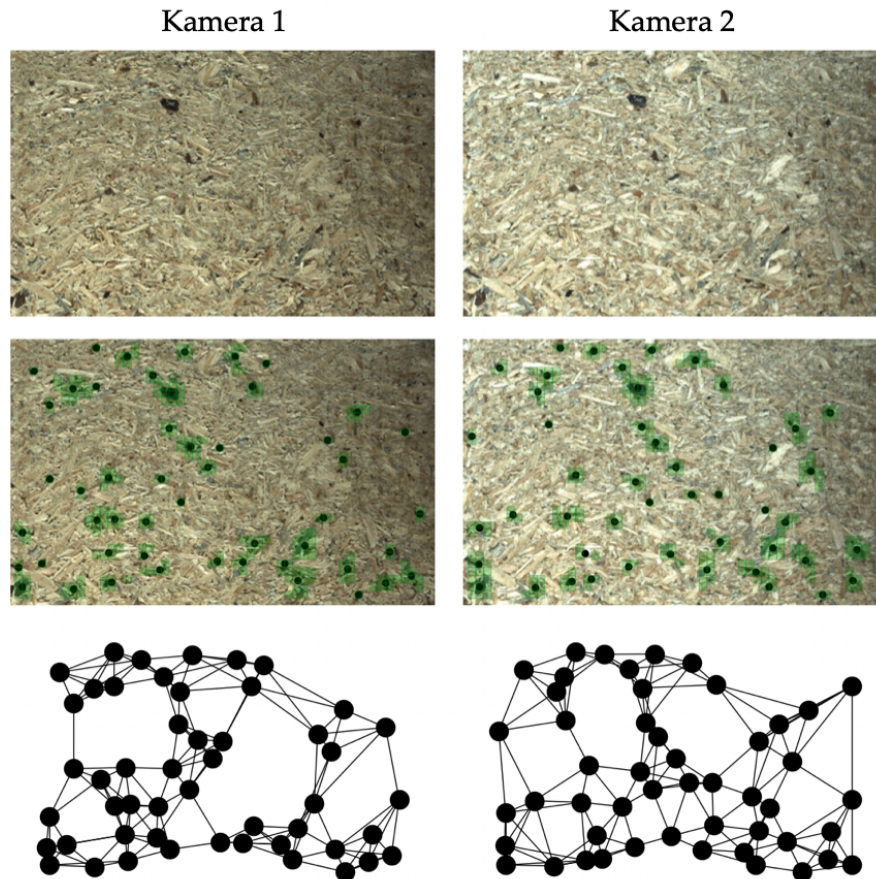
### 6.2.1 Evaluation von Szenario I

Szenario I ist nach Datensatznutzung unterteilt und fokussiert sich auf die Ergebnisse der ausgewählten Modelle. Für die Evaluation der Datensätze selbst wird auf Kapitel 6.1 verwiesen.

#### 6.2.1.1 Ergebnisse auf Datensatz A

In der Anwendung auf Datensatz A wurden bei der Anwendung des PCB Modells Prädiktionsgenauigkeiten von über 99 % erreicht (siehe Tab. 6.1). So

liegt bereits die Rank-1-Accuracy bei 99,86 % und steigt über die kommenden Ränge marginal weiter an. Die Standardabweichung der Ergebnisse, liegt zwischen 0,08 % und 0,1 % und ist damit verhältnismäßig niedrig.



**Abb. 6.1** Visualisierung der entscheidenden Features eines Klotzes aus Datensatz A mittels Graphdarstellung.

Im Kontrast dazu ist das deutlich einfachere Graphmodell in seinen Prädiktionen weniger akkurat. Die Rank-1-Accuracy des Modells liegt hier lediglich bei 53,8 %. Die Genauigkeit steigt bis zum zehnten Rang auf 88,85 %. Die Standardabweichungen sind im Vergleich zum PCB Modell deutlich höher. Dennoch liegt das Modell mit seinen Ergebnissen weit über der Baseline. Das Graphmodell ist darüber hinaus in seiner Funktionsweise vergleichsweise transparent, da die Knoten des Graphen visualisierbar sind und somit die Entscheidungsfindung des Modells repräsentieren. Ein Beispiel einer solchen Visualisierung ist in Abb. 6.1 dargestellt.

**Tabelle 6.1** Ranked Accuracy der Modelle, trainiert und getestet auf Datensatz A.

Modell	Top-k-Accuracy [%]				
	Rank 1	Rank 2	Rank 3	Rank 5	Rank 10
PCB	99,86 ± 0,1	99,9 ± 0,09	99,92 ± 0,1	99,94 ± 0,08	99,94 ± 0,08
Graph	53,8 ± 3,02	65,73 ± 2,28	73,1 ± 2,04	80,09 ± 2,1	88,85 ± 1,36
Ensemble	95,01 ± 0,48	97,01 ± 0,31	97,79 ± 0,4	98,37 ± 0,37	98,95 ± 0,39

Das Ensemblemodell liegt in seiner Genauigkeit zwischen den beiden vorherigen Modellen, mit einer Rank-1-Accuracy von 95,01 %. Die Prädiktionsgenauigkeit des Modells steigt bis zum zehnten Rang auf 98,95 % an und ist angesichts seiner simplen Architektur verhältnismäßig akkurat. Des Weiteren sind die Standardabweichungen des Modells deutlich geringer, als die des Graphmodells.

### 6.2.1.2 Ergebnisse auf Datensatz B

Für Datensatz B erzielte das PCB Modell auf Rang 1 eine Prädiktionsgenauigkeit von 92,63 % (siehe Tab. 6.2). Die Genauigkeit steigt bis Rang 10 auf 99,93 % an. Unter Nutzung von Datensatz B sind ähnliche Standardabweichungen zu beobachten, wie unter Nutzung von Datensatz A. Das Modell ist auf diesem Datensatz außerdem weniger performant, was womöglich am deutlich größeren Umfang und der Diversität des Datensatzes liegt. Dies ist insbesondere im ersten Rang sichtbar, nach welchem die Prädiktionsgüte bedeutend ansteigt.

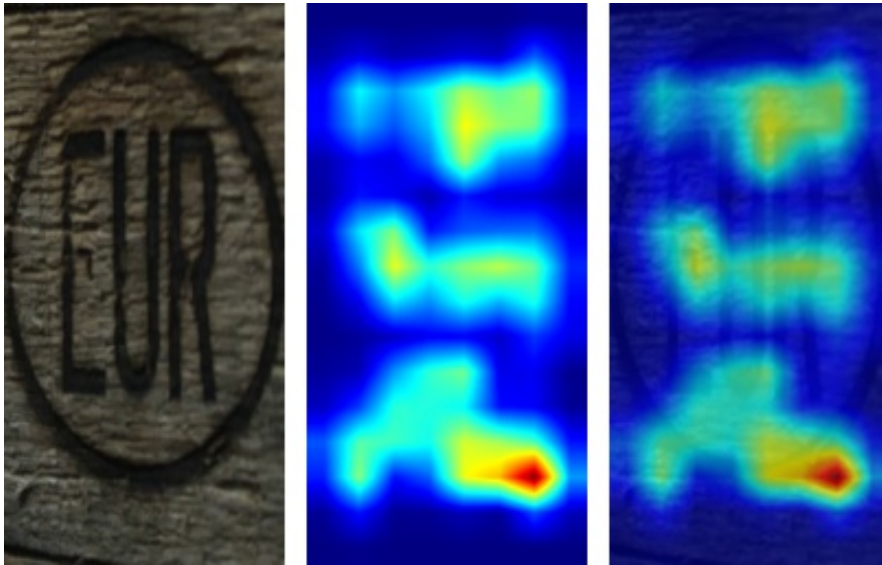
**Tabelle 6.2** Ranked Accuracy der Modelle, trainiert und getestet auf Datensatz B.

Modell	Top-k-Accuracy [%]				
	Rank 1	Rank 2	Rank 3	Rank 5	Rank 10
PCB	92,63 ± 0,4	98,37 ± 0,21	99,67 ± 0,09	99,93 ± 0,06	99,93 ± 0,05
Graph	47,64 ± 2,46	59,01 ± 2,24	65,59 ± 2,45	72,25 ± 2,21	80,93 ± 2,12
Ensemble	70,11 ± 3,41	77,7 ± 3,11	80,62 ± 2,68	84,37 ± 2,3	88,41 ± 1,94

Das Graphmodell erreicht auf Datensatz B eine Rank-1-Accuracy von 47,64 %. Die Performance des Modells steigt bis zum zehnten Rang auf bis zu 80,93 % an. Im Vergleich zu Datensatz A, scheint das Graphmodell auf Datensatz B leicht inakkurater zu sein. Als rudimentäres Modell weist das Graphmodell bei der Nutzung der Datensätze A und B somit vielversprechende Ergebnisse auf. Beim Training des Modells auf diesem Datensatz

wurde wiederholt das in einer Vorarbeit [KRM24] beschriebene Phänomen der korrelierten Features beobachtet.

Das Ensemblemodell erreicht auf diesem Datensatz eine Rank-1-Accuracy von 70,11 %. Die Ergebnisse steigen bis zum zehnten Rang auf eine Genauigkeit von 88,41 % an. Es sind außerdem ausgeprägte Standardabweichungen von bis zu 3,41 % zu beobachten. Das Ensemblemodell ist im Vergleich zu Datensatz A deutlich weniger performant auf Datensatz B. Insbesondere im Vergleich zum PCB Modell, scheint das Ensemblemodell somit weniger gut auf größere Datensätze zu skalieren.



**Abb. 6.2** Visualisierung der entscheidenden Features eines Klotzes aus Datensatz B mittels Activation Heatmap unter Nutzung des PCB Modells.

Für die zusätzliche Auswertung wurde bei diesem Versuch die Visualisierung der entscheidenden Features mittels Attention Heatmaps in Abb. 6.2 dargestellt. Analog zu Abb. 3.1 sind die signifikanten Bildabschnitte, aufgrund der Funktionsweise des PCB Modells, in Form von horizontalen Streifen zu beobachten.

### 6.2.1.3 Ergebnisse auf Datensatz C

In der Anwendung auf Datensatz C wurden bei der Anwendung des PCB Modells bereits auf dem ersten Rang Prädiktionsgenauigkeiten von 95,26 % erreicht (siehe Tab. 6.3). Über die kommenden Ränge steigt die Prädiktionsgenauigkeit auf bis zu 99,16 % an. Die dabei verzeichneten Standardabweichungen liegen unter 2 %.

Das Graphmodell erreicht auf Datensatz C lediglich eine Rank-1-Accuracy von 20,87 %. Die Prädiktionsgenauigkeit steigt auf dem zehnten Rang auf 54,15 % an. Aufgrund der gänzlich anderen Oberflächenbeschaffenheit des galvanisierten Stahls scheint der anomaliebasierte Ansatz des Graphmodells nur eingeschränkt für eine verlässliche Identifikation geeignet zu sein. Durch die deutlich eingeschränkte Genauigkeit des Modells lässt es sich nur als Teil eines Ensembles sinnvoll verwenden.

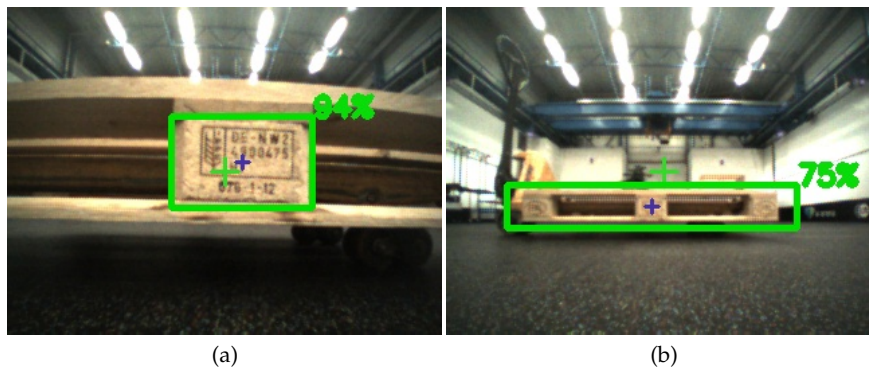
**Tabelle 6.3** Ranked Accuracy der Modelle, trainiert und getestet auf Datensatz C.

Modell	Top-k-Accuracy [%]				
	Rank 1	Rank 2	Rank 3	Rank 5	Rank 10
PCB	95,26 ± 1,91	96,78 ± 1,47	97,38 ± 1,1	98,26 ± 0,77	99,16 ± 0,22
Graph	20,87 ± 0,51	29,17 ± 0,53	34,61 ± 0,65	42,59 ± 1,43	54,15 ± 1,67
Ensemble	74,85 ± 2,79	81,31 ± 3,93	84,42 ± 4,23	87,55 ± 4,36	91,23 ± 4,01

Die Ergebnisse des Ensemblemodells ähneln der Performance, die auf Datensatz B erzielt wurde. Die Rank-1-Accuracy liegt hier bei 74,85 %. Anschließend steigt die Prädiktionsgenauigkeit des Modells auf bis zu 91,23 % an. Auch die beobachteten Standardabweichungen von bis zu 4,36 % ähneln den Ergebnissen unter Nutzung von Datensatz B.

### 6.2.2 Evaluation von Szenario II

Für das Training der zur Detektion genutzten YOLO Modelle wurde in allen vier Fällen eine  $IoU$  von über 0,8 erreicht (Modell 1a  $IoU_{max} = 0,85$ , Modell 1b  $IoU_{max} = 0,89$  und Modell 2a  $IoU_{max} = 0,87$ , Modell 2b  $IoU_{max} = 0,89$ ). Die Gewichtungen der performantesten Epochen wurden gespeichert und für die weiteren Versuche verwendet.



**Abb. 6.3** Visualisierung der Prädiktionen der Bounding Boxes zur Detektion der Klötze (a) und Paletten (b) mittels YOLO.

Basierend auf der Nutzung dieser Modelle wurde die im vorangehenden Kapitel beschriebene Mission 280 Mal erfolgreich durchgeführt. Ein Beispiel der Prädiktionen der YOLO Modelle ist Abb. 6.3 zu entnehmen. Die durchschnittliche Flugzeit betrug 59,37 s, die mittlere Flugzeit betrug 58,97 s. Der längste Flug dauerte 127,56 s und der kürzeste Flug wiederum 36,22 s. Die Flugzeit hing bei den Experimenten besonders von der Geschwindigkeit ab, mit der die Bilder der Klötze aufgenommen wurden. Sofern die dafür festgelegten Konditionen hinsichtlich der Bounding Box Größe nicht erreicht wurden, verlängerte dies die Flugzeit bedeutend.

In Abb. 6.4 ist die Annäherung der Drohne zur Bilderstellung des mittleren Klotzes dargestellt. In diesem Zustand würde die Drohne einen Großteil der Flugzeit verweilen, wobei die Regelung nur kleinste Steuersignale sendet. Auf diese Weise wird versucht, eine geeignete Bounding Box Größe zu erreichen und die Bilderstellung abzuschließen. In Summe betrug die Flugzeit 16.622,69 s und somit 4 Stunden und 37 Minuten. Die Dauer der Durchführung der Experimente (exklusive der Datensatzerstellung, des Modelltrainings und der Auswertungen) betrug circa 16 Stunden und 18 Minuten.



**Abb. 6.4** Die in Szenario II verwendete Drohne im Prozess der Bildaufnahme für die anschließende Identifikation der Palette.

Nachdem die drei erforderlichen Bilder pro Palette erstellt und an den Server geschickt worden sind, wurde die Identifikation der Palette initiiert. Noch vor der Landung der Drohne an ihrem Ausgangspunkt wurde die Berechnung durchgeführt. Im Durchschnitt fand die Komputation innerhalb von 248,66 ms, im Mittel innerhalb von 175 ms statt. Es sind regelmäßig Ausreißer (in 21 Fällen) bei der Berechnungszeit festzustellen. Dies sind die einzigen Fälle, in denen die Berechnungszeit länger als 250 ms betrug. In diesen Fällen

wich die Berechnungszeit deutlich von den sonstigen Zeiten ab und lag jeweils bei über eine Sekunde (Maximalwert von 1.338,15 ms). Diese Ausreißer erklären den bedeutenden Unterschied zwischen der durchschnittlichen und der mittleren Berechnungszeit. Die kürzeste Berechnungszeit wurde mit 115,39 ms beobachtet. Die Ergebnisse der Flugzeit und der Berechnungszeit sind Abb. 6.5 zu entnehmen.

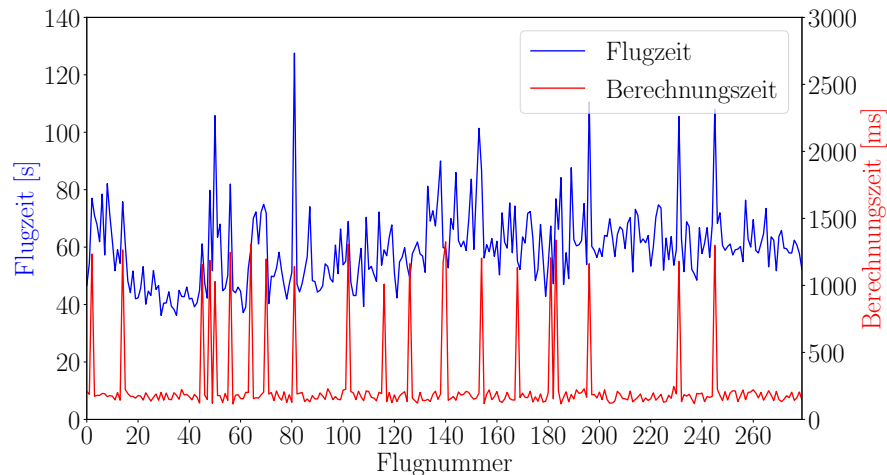


Abb. 6.5 Flugzeiten und Berechnungszeiten aller 280 Flüge für Szenario II.

Der Neuheitsschwellwert (in Abb. 6.6 als gestrichelte Linie dargestellt) wurde bei einer Distanz von 1,5 festgelegt. Dieser Wert wurde in drei der 280 Flüge überschritten, womit eine Palette fälschlicherweise als neu identifiziert worden wäre. Im Durchschnitt liegt die minimale Distanz bei 0,99 und die durchschnittliche Distanz bei 1,69. Es wurde deshalb überwiegend eine bedeutende Distanz zwischen der zugeordneten ID und den restlichen IDs festgestellt, was eine korrekte Zuordnung gewährleistet. Der durchschnittliche Wert liegt ebenfalls über dem festgelegten Neuheitsschwellwert. Die Werte pro Flug sind Abb. 6.6 zu entnehmen.

Des Weiteren wurden 19 Flüge (6,35 % der Flüge) wiederholt, da es nicht zu einer Identifikation kam. Dies lag hauptsächlich an vorgeschalteten Komplikationen, wie der Steuerung der Drohne (ROS, MQTT) oder Hardwareproblemen. In nur vier Fällen hat die eigentliche Regelung der Drohne (im Sinne der PID-Regelung und YOLO Modellnutzung) zu einer Komplikation geführt, die einen Abbruch des Flugs verursachte. In weiteren sechs Fällen haben die Batterien die Rotoren der Drohne nicht mit der nötigen Spannung versorgt, die für eine erfolgreiche Durchführung des jeweiligen Flugs nötig gewesen wäre. Dies führte in der Regel zu Abstürzen kurz nach dem Abheben der Drohne.

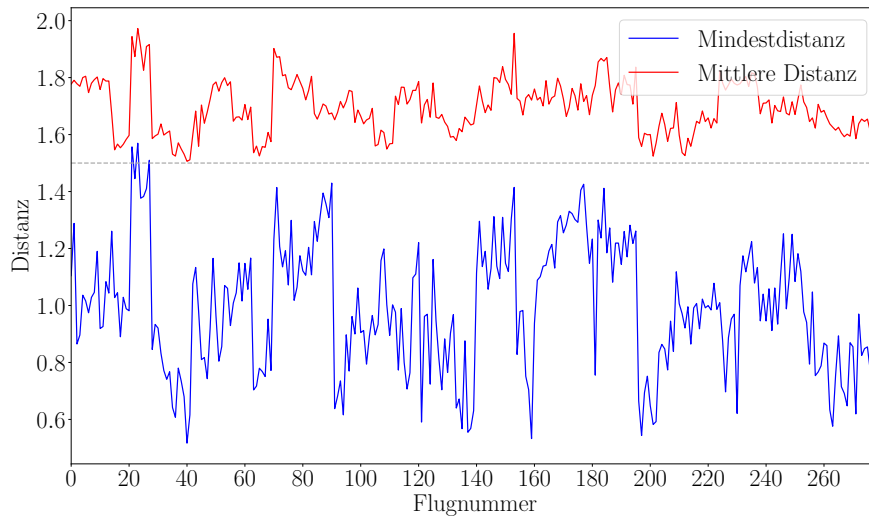


Abb. 6.6 Minimale und mittlere Embedding Distanzen der 280 Flüge für Szenario II.

In den übrigen neun Fällen traten Verbindungsprobleme zwischen der Drohne und dem Server auf, durch die nur noch die Kommunikation von Drohne zu Server (unidirektional) aufrechterhalten wurde. Die Drohne wurde somit nicht mehr mit Steuersignalen versorgt und verblieb bis zur manuellen Landung an ihrem letzten Aufenthaltsort. Von den 280 gelungenen Flügen fand in 92,14 % der Flüge (258 Flüge) eine erfolgreiche Identifikation der Palette statt, im Kontrast zu einer Baseline Accuracy von 2,5 %. Unter Betrachtung der zweitniedrigsten Distanz ist eine Rank-2-Accuracy von 97,85 % festzustellen, was 274 erfolgreichen Identifikationen entspräche. Die Übersicht der Flüge und ihrer Ergebnisse sind Abb. 6.7 zu entnehmen.

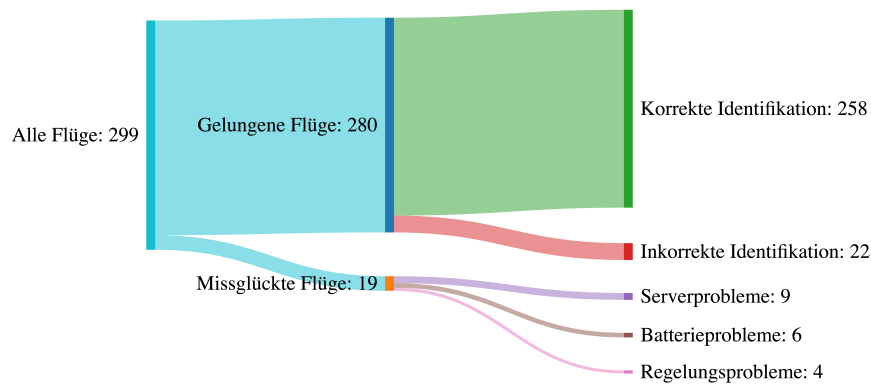


Abb. 6.7 Sankey Diagramm der 280 gelungenen und 19 gescheiterten Flüge für Szenario II.

Beispiele der erstellten Bilder einer Palette (ID 19) und der zuvor bereitgestellten Ground Truth der ID sind Abb. 6.8 zu entnehmen. Diese Abbildung stellt ein Positivbeispiel von sowohl Detektion als auch Identifikation dar. Die aufgezeichneten Bilder sind scharf, der Klotz zentriert und es ist weder zu viel Hintergrund aufgenommen worden, noch ein relevanter Teil des Bildes abgeschnitten worden.



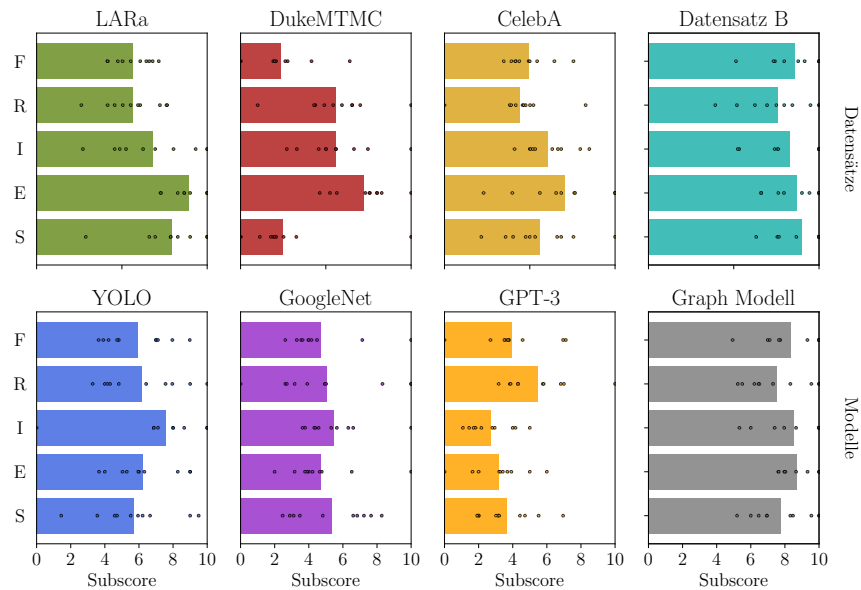
**Abb. 6.8** Beispiel der in Szenario II erfassten Bilder, skaliert auf eine 3:4 Aspect Ratio (ID 19, Klotz 3, Ground Truth oben links, Flüge 1 bis 7 von oben links bis unten rechts).

### 6.3 Verlässlichkeitsuntersuchung

In diesem Kapitel werden die durchgeführten Untersuchungen hinsichtlich der Verlässlichkeitsaspekte aus Kapitel 5 betrachtet. Dabei wird die Auswertung der Bewertung der Verlässlichkeit pro Aspekt und als Ganzes präsentiert. Die entwickelte Bewertungsmethode wird weiterhin hinsichtlich ihrer Tauglichkeit evaluiert. Darüber hinaus werden Untersuchungen des Identifikationsprozesses vorgestellt, welche die Ergebnisse aus dem vorangegangenen Abschnitt hinsichtlich ihrer Verlässlichkeit ergänzen.

Um die Anwendbarkeit und Reliabilität des präsentierten Verfahrens zu untersuchen, wurden die bereits vorgestellten Modelle und Datensätze von zehn Nutzern hinsichtlich ihrer Verlässlichkeit bewertet. Die resultierenden Bewertungen der Nutzer für jeden einzelnen Verlässlichkeitsaspekt sind in Abb. 6.9 zu sehen. In einigen Fällen sind hohe Varianzen zwischen den erhaltenen Punktzahlen pro Aspekt festzustellen, so etwa im Fall von DukeMTMC und GPT-3. Für beispielsweise YOLO und GoogleNet, weichen die Ergebnisse je Nutzer jedoch nur unbedeutend voneinander ab. Auch pro Aspekt sind unterschiedliche Streuungsgrade zu beobachten. So sind etwa im Vergleich zum Aspekt der Sicherheit von YOLO, die Spannbreiten der Bewertungen für den Aspekt der Sicherheit von DukeMTMC, der Erklärbarkeit des Graphmodells, oder der Fairness von LARA deutlich geringer. In letzterem Fall ist die Bewertungen der zehn Nutzer über nahezu die gesamte Bewertungsskala verteilt. Bis auf diesen Fall sind eingeschränkte Spannbreiten der Bewertungs-

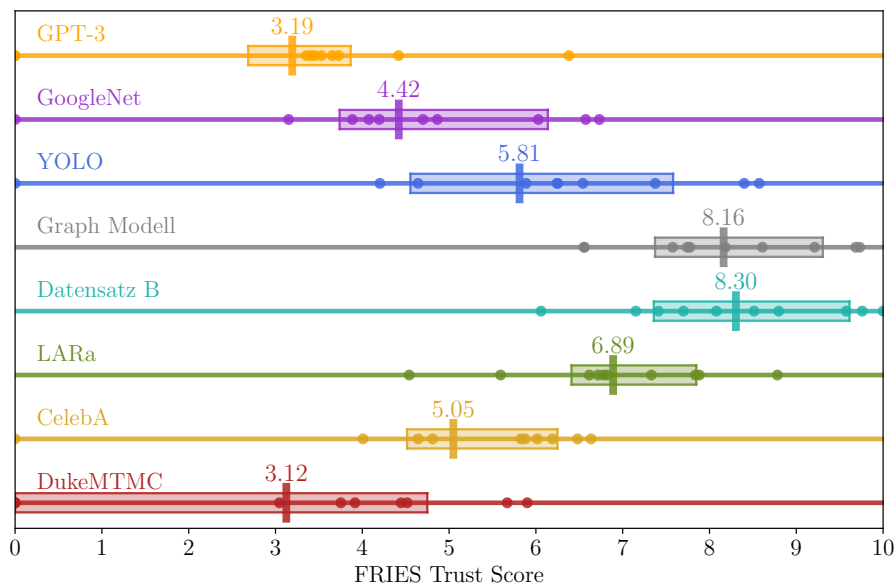
streuung zu beobachten, die teilweise einzelne Ausreißer aufweisen (siehe etwa den Aspekt der Sicherheit von LARa, den Aspekt der Robustheit von CelebA, oder den Aspekt der Fairness von GoogleNet). Die Erklärbarkeit ist bei drei der vier Datensätze der am höchsten bewertete Aspekt. Im Fall von Datensatz B liegt die Bewertung der Erklärbarkeit dennoch nicht weit gegenüber der Bewertung der Sicherheit zurück. Auch die Bewertungen der Aspekte der Robustheit und Integrität sind für die Datensätze ähnlich ausgefallen. Ein ähnliches Muster lässt sich für die untersuchten Modelle nicht beobachten.



**Abb. 6.9** Die Werte der Aspekte pro Modell und Datensatz, basierend auf den Bewertungen von zehn Befragungsteilnehmern.

Die Verlässlichkeitswerte der Modelle und Datensätze in ihrer Gesamtheit sind in Abb. 6.10 dargestellt. Die Mittelwerte sind in der Grafik als vertikale Balken dargestellt, die Rechtecke um sie herum repräsentieren die 20 %- und 80 %-Quintile. Wie in Abb. 6.9 dargestellt, bewerteten die Nutzer die Aspekte der Modelle und Datensätze in den meisten Fällen ähnlich. Es sind gewisse Streuungen zu beobachten, die je nach Modell und Datensatz unterschiedlich bedeutend ausfielen. Die Streuung der Ergebnisse ist bei GPT-3 am niedrigsten. Dies ist möglicherweise auf seine aktuell ausgeprägte Präsenz in Forschung und Medien zurückzuführen, wodurch das Modell allen Nutzern sehr gut bekannt ist. Die größte Streuung innerhalb der Ergebnisse ist bei DukeMTMC zu beobachten. Hier beginnt das 20 %-Quartil bereits mit einem Wert von 0, da mehrere Nutzer die Verlässlichkeit des Datensatzes

in Gänze mit einem Verlässlichkeitswert von 0 bewertet haben. Da jedoch einige Nutzer die Verlässlichkeit von DukeMTMC als mittelmäßig interpretierten (Verlässlichkeitswerte mittig auf der Skala), ergibt sich somit eine verhältnismäßig hohe Varianz. Wie eingangs beschrieben, handelt es sich bei DukeMTMC um einen Datensatz, der in der Forschungsgemeinde als problematisch gilt und aufgrund von Datenschutzverletzungen zurückgezogen wurde. In Anbetracht des faktisch fragwürdigen Hintergrunds des Datensatzes war für DukeMTMC eine niedrige Bewertung auf der Verlässlichkeitsskala zu erwarten. Dennoch bewerteten zwei der Nutzer den Datensatz mit einem Wert von circa 5 und der Datensatz erhält somit im Mittelwert einen Verlässlichkeitswert von 3,12. Dies ist möglicherweise aufgrund einer missverständlichen Interpretation der Sachlage oder einer einschränkbareren Nutzbarkeit des erstellten Bewertungsverfahrens der Fall. Nichtsdestotrotz erhielt der Datensatz den niedrigsten Verlässlichkeitswert von allen Modellen und Datensätzen.



**Abb. 6.10** Verlässlichkeitswerte pro Modell und Datensatz, basierend auf den Bewertungen von zehn Befragungsteilnehmern.

Der Datensatz LARa grenzt sich in seiner Bewertung von DukeMTMC und CelebA mit einem Wert von 6,89 deutlich ab. Datensatz B wurde als der vertrauenswürdigste Datensatz bewertet, mit einem Wert von 8,3. Das am verlässlichsten bewertete Modell war das Graphmodell, mit einem Wert von 8,16. Auch hier ist eine deutliche Abweichung von der Bewertung der restlichen Modelle festzustellen. Die hohen Verlässlichkeitswerte für die

drei am Standort Dortmund entwickelten Modelle und Datensätze weisen dennoch womöglich auf Biases der Nutzer hin.

Durch die Auswertungen zeigen sich insgesamt Bewertungsmuster und somit begrenzte Streuungsgrade zwischen den Bewertungen der Teilnehmer. Bei der Befragung der Nutzer und den vorangegangenen Untersuchungen wurde auch die Widersprüchlichkeit einiger Anforderungen an die Verlässlichkeit deutlich. Ein solches Beispiel ist die erhöhte Erklärbarkeit eines öffentlich zugänglichen Modells gegenüber der erhöhten Sicherheit, die durch die Sicherung eines proprietären Modells gewonnen wird.

Nach der Bewertung gaben die Nutzer Feedback zu dem verwendeten Verfahren. Es wurde darauf hingewiesen, die Bedeutung der OSD-Tabelle hätte deutlicher erklärt werden können, da einige Nutzer nicht mit dem FMEA Verfahren vertraut waren. Die teils positive und teils negative Formulierung der Bewertungsfragen ist ebenfalls als eine Herausforderung im Bewertungsprozess genannt worden. Es wurde angemerkt, die aufgelisteten Risiken hätten expliziter erläutert werden können, wodurch eine erhöhte Eindeutigkeit hätte erreicht werden können. Dies wäre ebenfalls durch eine exemplarische Bewertung eines Modells oder eines Datensatzes zur Orientierung umsetzbar, was von einem der Nutzer angemerkt worden ist. Während die Werte 0 und 10 von den Nutzern als eindeutig bezeichnet wurden, wurde angemerkt, vor allem der mittlere Bereich der Skala sei mehrdeutig. Es sei demnach schwierig für einen Nutzer gewesen, den Unterschied zwischen der Bedeutung einer Bewertung von 4 oder 5 festzumachen. Letztlich wird in diesem Rahmen die Auffassung von [Dod+19] bekräftigt, nach welcher die Sicherstellung von Fairness (und somit der Verlässlichkeit insgesamt) ein Prozess sei, der den Menschen als Beurteilungskomponente weitergehend beinhaltet („Human in the Loop“). Diese Auffassung wird nicht zuletzt durch den Mangel algorithmischer Alternativen in der relevanten Literatur unterstrichen. Folglich wird Subjektivität auf absehbare Zeit ein Teil des Operationalisierungsprozesses der Verlässlichkeit bleiben.

## 6.4 Erkenntnisgewinn und kritische Würdigung der Arbeit

Basierend auf den erarbeiteten Ergebnissen wird in diesem Kapitel eine kritische Reflexion des Erkenntnisgewinns dieser Arbeit durchgeführt. Dabei wird in einem ersten Schritt Bezug zur wissenschaftstheoretischen Positionierung der Arbeit genommen, welche die Ansprüche an die Ergebnisse dieser Arbeit definiert. Es wurde ein für Anwender zugängliches, Open Source Verfahren zur Identifikation logistischer Entitäten erarbeitet, das Identifikationsgenauigkeiten von über 90 % ermöglicht. Diese Genauigkeit wurde auch bei der Nutzung in einem industrieähnlichen Umfeld erzielt. Durch die genutzte Versuchsplanung sind die Ergebnisgenauigkeiten mit einer Konfidenz

von 95 % angegeben. Die verwendete Hardware ist ebenfalls zugänglich, ökonomisch nicht restriktiv und stellt nur eine von diversen Möglichkeiten dar, da eine erfolgreiche Replikation der vorgestellten Ergebnisse nicht von einer spezifischen Hardware abhängt. Die vorgestellten Verfahren bieten sich somit als Teil eines Lagers der Zukunft an, in welchem datengetriebene Logistik und Computer Vision Methoden eine entscheidende Rolle spielen werden.

Dennoch weist die Identifikation logistischer Entitäten einige Herausforderungen auf. So spielen je nach Modell etwa Einschränkungen in der Identifikationsgenauigkeit und gegebenenfalls auch Umgebungsbedingungen, eine tragende Rolle hinsichtlich der industriellen Umsetzbarkeit des hier vorgestellten Vorgehens. Die industrielle Umsetzung der in Szenario II präsentierten mobilen Lösung bedarf der Nutzung eines vollumfänglichen Motion Capture Systems zur Navigation der Drohne, das kostenintensiv und in vielen Lagerumgebungen nur eingeschränkt umzusetzen ist. Die Nutzung der vorgestellten Verfahren ist ebenfalls eingeschränkt auf Entitäten, die über einzigartige Oberflächenmerkmale verfügen. Über die Skalierbarkeit in einem industriellen Szenario ist basierend auf den vorliegenden Ergebnissen noch keine Aussage zu treffen, da in der Industrie Millionen von Entitäten miteinander zu vergleichen sind. Die ökonomische Sinnhaftigkeit, gilt es je nach Anwendung zu überprüfen und abzuwägen. Diese steht im direkten Vergleich zu alternativen Methoden, wie etwa der Nutzung von RFID oder QR-Codes. Im Falle der Identifikation von Ladungsträgern stellt zuletzt auch die Verheiratung des Guts und der Palette selbst eine mögliche Anschlussforschung dar.

Das entwickelte Bewertungsverfahren für das Konzept der Verlässlichkeit im Kontext der Identifikation und des Maschinellen Lernens im weiteren Sinne basiert auf einer vorangehenden Recherche und Definition der relevantesten Aspekte der Verlässlichkeit. Durch die Identifikation und Definition jener Aspekte ist ein trennscharfes Verständnis und damit eine Definition für den Begriff der Verlässlichkeit beigetragen worden. Auf dieser Basis wird es der Forschungsgemeinde ermöglicht, Modelle und Datensätze im Kontext des Maschinellen Lernens miteinander hinsichtlich ihrer Verlässlichkeit zu vergleichen. Das entwickelte Bewertungsverfahren spielt für diese Komparabilität eine entscheidende Rolle und zeigt unter erster Anwendung reliable Ergebnisse.

Dennoch sind das entwickelte Bewertungsverfahren und die erarbeiteten Definitionen weiter ausbaufähig. Sowohl die zu der Erarbeitung der Aspekte der Verlässlichkeit führende Kategorienbildung, als auch die Erstellung des Bewertungsverfahrens, sind in Zukunft von der Forschungsgemeinde einzusetzen und zu beurteilen, um eine fundierte Aussage über ihren Nutzen zu ermöglichen. In den noch eingeschränkten Anwendungen der Bewertung wurden weiterhin gewisse Streuungen der Ergebnisse beobachtet. Die

Anzahl der durchgeführten Bewertungen ist eingeschränkter Natur und ist zukünftig um eine umfangreichere Studie zu erweitern, um eine statistische Aussagekraft zu erreichen. Darüber hinaus widersprechen sich einige Aspekte der Verlässlichkeit in ihren Anforderungen gegenseitig, was die Kalkulation des Verlässlichkeitswerts beeinflusst. Zuletzt bleibt durch die Subjektivität der Bewertung, die Einschätzung der Verlässlichkeit von Modellen und Datensätzen aktuell ein HIL-Prozess. Ob eine vollständig objektive und automatisierte Bewertung möglich wäre, bleibt weiterhin offen.

In diesem Kontext und in Anbetracht der im vorangegangenen Kapitel präsentierten Ergebnisse lassen sich die in Kapitel 1 formulierten Forschungsfragen nun wie folgt beantworten:

FF1 *Wie können logistische Entitäten anhand ihrer inhärenten visuellen Merkmale identifiziert werden?*

↔ Logistische Entitäten, die uneinheitliche Oberflächenmerkmale aufweisen, können anhand der Einzigartigkeit ebenjener Merkmale visuell identifiziert werden. Dafür können gängige Identifikationsalgorithmen sowie weniger komplexe Ensemblemodelle verwendet werden, sofern im Vorfeld geeignete Datensätze der jeweiligen Entitäten erstellt worden sind. In Abhängigkeit der Komplexität des genutzten Modells und Datensatzes werden auch in industrieähnlichen Umgebungen Identifikationsgenauigkeiten von über 90 % erzielt. Die mobile, autonome und aufwandsarme Umsetzbarkeit des Identifikationsverfahrens ist demonstriert worden.

FF2 *Wie kann der Begriff der Verlässlichkeit im Kontext der Identifikation definiert und quantifiziert werden?*

↔ Basierend auf einer induktiven Kategorienbildung wurde der Begriff der Verlässlichkeit über seine relevanten Aspekte der Fairness, Robustheit, Integrität, Erklärbarkeit und Sicherheit operationell definiert. Konzeptionell umfasst der Begriff der Verlässlichkeit den fairen Umgang mit Daten, die robuste Leistungserbringung beim Auftreten anomaler Daten, die Sicherstellung der Integrität von Daten und Modellen, die Bereitstellung erklärbarer Entscheidungen sowie die sichere Nutzung vertraulicher Informationen. Eine Quantifikation der Verlässlichkeit für Datensätze und Modelle des Maschinellen Lernens ist mittels der in dieser Arbeit vorgestellten FRIES-Bewertungsmethode möglich und attribuiert den Modellen und Datensätzen Verlässlichkeitswerte (FRIES Trust Score) auf einer Skala von 0 bis 10. Eine erste Untersuchung der Reliabilität dieser Bewertungsmethode wurde ebenfalls durchgeführt.



## Kapitel 7

# Zusammenfassung und Ausblick

„The game of science is, in principle, without end. He who decides one day that scientific statements do not call for any further test, and that they can be regarded as finally verified, retires from the game.“

– Karl R. Popper, „Logik der Forschung“, Kapitel XI (1959)

Im Rahmen dieser Arbeit wurde die Identifikation logistischer Entitäten unter Verwendung ihrer inhärenten visuellen Merkmale untersucht. Es wurden dafür Experimente durchgeführt, mithilfe derer die industrielle Anwendung des erarbeiteten Verfahrens getestet und evaluiert wurde. Dabei wurde ein besonderer Schwerpunkt auf die Verlässlichkeit des Verfahrens gelegt, für die eine Definition und eine darauf basierende Metrik erstellt wurden.

### 7.1 Erkenntnisse und wissenschaftlicher Beitrag

Zusammenfassend lassen sich folgende Schlüsselbeiträge und Erkenntnisse der Arbeit formulieren:

- Es wurden neue, hochwertige und industriell nutzbare Datensätze für die Identifikation logistischer Entitäten erstellt und für die Forschungsgemeinschaft öffentlich zur Verfügung gestellt. Die Nutzbarkeit der Datensätze wurde im Rahmen dieser Arbeit evaluiert, indem sie als Trainingsdaten für die genutzten Modelle dienten.
- Es wurde erstmals ein Modell zur Identifikation von Personen für die Identifikation logistischer Entitäten umfunktioniert. Es wurde darüber hinaus ein neues Graphmodell erarbeitet, das spezifisch für die Identifikation logistischer Entitäten mit einzigartigen, visuellen Oberflächenmerkmalen entwickelt wurde. In Abhängigkeit des genutzten Modells wurde eine Identifikationsgenauigkeit von bis zu 99,86 % erreicht. Es wurde der Mehrwert der Nutzung rudimentärer, heterogener Ensemblemodelle festgestellt, die auf den verwendeten Datensätzen verhältnismäßig hohe Prädiktionsgenauigkeiten von bis zu 95,01 % aufweisen.

- Es wurde somit ein Proof of Concept hinsichtlich der Identifikation logistischer Entitäten, basierend auf ihren inhärenten, visuellen Merkmalen erbracht. Die verwendeten Entitäten bestehen aus Holz und Metall und weisen einzigartige Oberflächenmerkmale auf, spezifisch die Holzmaserung und das Muster des Pressspans der Palettenklötze sowie die Zinkblumen des galvanisierten Stahls der Stapeltransportkisten. Diese logistischen Entitäten stehen damit stellvertretend für andere Entitäten ähnlicher Beschaffenheit. Sie dienen als Inspiration für die Herstellung neuer Entitäten, die eine eindeutige Identifikation zulassen. Ein Beispiel dafür stellt die Spritzgussherstellung einer Entität aus Polymer dar, wobei mehrfarbiges Granulat verwendet wird. Auf diese Weise entstehen einzigartige Oberflächenmuster, die wie in dieser Arbeit beschrieben verwendbar wären. Eine geeignete logistische Entität wäre dafür der weit verbreitete Kleinladungsträger, der aktuell einfarbig hergestellt wird.
- Es wurde eine empirische Evaluation des Identifikationsverfahrens in Szenario II durchgeführt. Eine Drohne wurde eingesetzt, die automatisch Paletten anfliegt, ihre Klötze erfasst und aufnimmt. Sie nutzt dafür eine Kombination aus einer YOLO-basierter Objekt Detektion und einem PID-Regler. Mit einer Konfidenz von 95 % wurde eine Identifikationsgenauigkeit von 92,14 % ermittelt.
- Es wurde eine neue und für die Operationalisierung nutzbare Definition des Begriffs der Verlässlichkeit im Maschinellen Lernen vorgestellt. Dieser Definition zugrunde liegen die Definitionen der mittels Literaturrecherche identifizierten Teilaspekte (Fairness, Robustheit, Integrität, Erklärbarkeit, Sicherheit) der Verlässlichkeit. Basierend auf der Literaturrecherche wurde eine induktive Kategorienbildung der genannten Aspekte durchgeführt. Die Definition dient als Vorschlag für eine einheitliche Nutzbarkeit und Vergleichbarkeit in der Forschungsgemeinde, die bislang fehlt.
- Aufbauend auf der Definition der Verlässlichkeit, wurde eine Verlässlichkeitsmetrik erarbeitet. Unter Verwendung dieser Metrik wird es Nutzern ermöglicht, Modelle und Datensätze aus dem Maschinellen Lernens hinsichtlich ihrer Verlässlichkeit und ihrer zugrundeliegenden Aspekte zu bewerten. Die Reliabilität der Metrik wurde mittels des Vergleichs der Bewertung von 10 Nutzern aus dem Bereich der Anwendung und der Forschung der Maschinellen Lernmethoden untersucht. In diesem Rahmen wurde die Nutzbarkeit der Metrik als erster Schritt der Operationalisierung festgestellt, wobei in Abhängigkeit der Modelle und Datensätze zum Teil ausgeprägte Streuungen der erzielten Ergebnisse vorliegen.

## 7.2 Forschungsausblick

Die Grenzen und Einschränkungen der Ergebnisse dieser Arbeit wiederum ebnen den Weg für Anschlussforschung. Diese Ergebnisse ermöglichen demnach die Weiterführung der genannten Vorarbeiten, die unter anderem folgende Themen beinhalten:

- Da die Erstellung von Trainingsdatensätzen einen hohen Aufwand mit sich bringt, ist die Erstellung synthetischer Datensätze von Interesse. Im Kontext der industriellen Nutzung ist die Erstellung von Datensätzen teilweise gänzlich ausgeschlossen. Dafür wurden in weiteren Vorarbeiten bereits erste Schritte unternommen, etwa die Erstellung mittels GAN (Generative Adversarial Network) generierter Bilder von Palettenklötzen für die Nutzung im Kontext der Re-identifikation [Rut+22b] oder auch für die Nutzung im Kontext der Objekt Detektion [Hol+23]. Diese ersten Ergebnisse gilt es weiterzuentwickeln und die Nutzung größerer, synthetischer Datensätze der Nutzung von Originaldaten gegenübergestellt werden.
- Die algorithmische Erarbeitung eines Schwellwerts zur Novelty Detection würde zu einer weiteren Verbesserung von Szenario II führen. Somit würden Open World Szenarien mit höherer Verlässlichkeit umgesetzt und, unter einer Annahme eines offenen Sets an Paletten, flexibler eingesetzt werden. Auch in diesem Kontext bestehen bereits Vorarbeiten [Sch+23], spezifisch eine erste algorithmische Lösung mittels Neighborhood Structure Aware Novelty Detection (NSAND).
- Eine der Limitierungen der visuellen Identifikation logistischer Entitäten ist die Alterung und dadurch visuelle Veränderung ebenjener. Dies gilt besonders bei Entitäten, die aus organischem Material bestehen. Durch solche visuellen Veränderungen würde die Genauigkeit der Identifikation über die Zeit hinweg sinken, was die Verlässlichkeit des Vorgehens beeinträchtigt. Aus diesem Grund ist in einer Vorarbeit die synthetische Alterung der Bilder von Palettenklötzen mittels GANs untersucht worden [Mat+23]. Aufbauend auf dieser Vorarbeit gilt es, die Alterung von realen Klötzen und den Einfluss auf die Identifikationsgenauigkeit weiter zu untersuchen.
- Neben der Umsetzung der Identifikation an Bord einer Drohne (Szenario II), bietet sich die Nutzung eines bodengebundenen Roboters an. Eine solche Umsetzung stellt eine Alternative dar, die höhere Hardwareressourcen zur Verfügung stellt, da die verwendete Drohne nur eine deutlich eingeschränkte Traglast ermöglicht. In einer Vorarbeit wurde eine Alternative an Bord eines DJI Robomasters vorgestellt. Durch die Nutzung des Robomasters ist ergänzend die Nutzung einer 5G-basierten Datenübertragung ermöglicht worden [Rut+23]. Die automatisierte Steuerung via 5G, sowie gänzlich andere Ansätze der mobilen Identifikation,

sind in diesem Kontext weiter von Interesse. Die Umsetzung einer SLAM (Simultaneous Localization and Mapping) Navigation, welche die Notwendigkeit eines Motion Capture Systems obsolet macht, ist in einem solchen Szenario vielversprechend.

- Die erzielten Genauigkeiten sind je nach Modell und Datensatz weiter optimierbar. Unter Nutzung des Graphmodells oder des Ensembles wäre eine Rank-1 Accuracy von über 90 % erstrebenswert. Dies ermöglicht eine verlässliche Implementierung eigens entwickelter, verlässlicher Methoden, auch unter Nutzung eingeschränkter Hardwareressourcen.
- Über die Skalierbarkeit der aktuellen Ergebnisse ist nur in begrenztem Maße eine Aussage zu treffen. Nichtsdestotrotz ist eine performante Identifikation gerade in industriellen Umgebungen, in denen Hunderttausende oder gar Millionen logistischer Entitäten voneinander unterschieden werden, von Interesse. Weitere Untersuchungen hinsichtlich der Skalierbarkeit und gegebenenfalls auch die Nutzung von Distraktorbildern, empfehlen sich demnach als Anschlussforschung.
- Die weitere Verwendung, Evaluierung und die Reliabilitätsuntersuchung der entwickelten Metrik und Definition der Verlässlichkeit stellt einen wichtigen Teil dieses Ausblicks dar. Insbesondere die Verwendung im industriellen Kontext und die Reliabilität der Nutzung der Metrik innerhalb einer Nutzergruppe spielt dabei eine entscheidende Rolle.
- Die Prozesssicherheit ist in der industriellen Anwendung von höchster Relevanz, da Fehler weitreichende Konsequenzen mit sich bringen. Die Untersuchung der Möglichkeit der Einbindung eines Menschen als Kontrollinstanz des Identifikationsprozesses ist demnach von Interesse. Darüber hinaus ist auch eine weitreichende Untersuchung der Akzeptanz des vorgeschlagenen Vorgehens in der Industrie von großer Relevanz.
- Die Mission der Drohne aus Szenario II um ein beidseitiges Abfliegen der respektiven Palette zu ergänzen, würde die Robustheit des Verfahrens weiter erhöhen, da mehr Bildinformation pro Palette verwendet werden würde. Die Nutzung von sechs Klötzen einer Palette in einem offline Szenario wurde bereits in einer Vorarbeit untersucht [Rut+22a] und unterstützt diese Annahme. Der kontrollierte und zuverlässige Flug um die Palette herum, stellt dabei jedoch eine Herausforderung seitens der Regelung der Drohne dar. In diesem Kontext stellt auch die Verheiratung des Guts und der Palette eine relevante Herausforderung dar, deren Lösung jedoch für die industrielle Nutzung von großer Bedeutung ist.

Insgesamt bieten die Ergebnisse dieser Arbeit somit viele weitere Forschungsansätze mit fundamentaler Relevanz für sowohl Industrie als auch Wissenschaft.

## Literaturverzeichnis

- [Ack90] Kenneth B. Ackerman. „Warehousing: Origins, History and Development“. In: *Practical Handbook of Warehousing*. Springer, 1990, S. 3–11. DOI: 10.1007/978-1-4757-1194-3\_1.
- [ACL05] Kiam H. Ang, Gregory Chong und Yun Li. „PID Control System Analysis, Design, and Technology“. In: *IEEE Transactions on Control Systems Technology* 13.4 (2005), S. 559–576. DOI: 10.1109/TCST.2005.847331.
- [AD21] Fatih Aksu und Cem Direkoğlu. „Lightweigth Convolutional Neural Networks for Person Re-Identification“. In: *3rd IEEE International Congress on Human-Computer Interaction, Optimization and Robotic Applications*. 2021, S. 1–5. DOI: 10.1109/HORA52670.2021.9461269.
- [Ada20] Henry Adams. „The Need for More Care in Warehouse Design.“ In: *Architecture, a Magazine of Architecture and the Applied Arts and Crafts* 13.3 (1920), S. 85–91.
- [Adj+20] Insaf Adjabi, Abdeldjalil Ouahabi, Amir Benzaoui und Abdelmalik Taleb-Ahmed. „Past, Present, and Future of Face Recognition: A Review“. In: *Electronics* 9.8 (2020), S. 1188. DOI: 10.3390/electronics9081188.
- [AG85] Jalal Ashayeri und Ludo F. Gelders. „Warehouse Design Optimization“. In: *European Journal of Operational Research* 21.3 (1985), S. 285–294. DOI: 10.1016/0377-2217(85)90149-3.
- [Ago+13] Federica Agosta, Sara Sala, Paola Valsasina, Alessandro Meani, Elisa Canu, Giuseppe Magnani, Stefano F. Cappa, Elisa Scola, Piero Qatto, Mark A. Horsfield, Andrea Falini, Giancarlo Comi und Massimo Filippi. „Brain Network Connectivity Assessed Using Graph Theory in Frontotemporal Dementia“. In: *Neurology* 81.2 (2013), S. 134–143. DOI: 10.1212 / WNL.0b013e31829a33f8.
- [AHM95] Peter Auer, Robert C. Holte und Wolfgang Maass. „Theory and Applications of Agnostic PAC-learning with Small Decision Trees“. In: *Machine Learning Proceedings*. Elsevier, 1995, S. 21–29. DOI: 10.1016/B978-1-55860-377-6.50012-8.
- [Aka+22] Chafic Abou Akar, Jimmy Tekli, Daniel Jess, Mario Khoury, Marc Kamradt und Michael Guthe. „Synthetic Object Recognition Dataset for Industries“. In: *2022 35th Conference on Graphics, Patterns and Images*. Bd. 1. 2022, S. 150–155. DOI: 10.1109/SIBGRAP155357.2022.9991784.

- [Arr+20] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier D. Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina und Richard Benjamins. „Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI“. In: *Information fusion* 58 (2020), S. 82–115. DOI: 10.1016/j.inffus.2019.12.012.
- [AS21] Pekka Ala-Pietilä und Nathalie A. Smuha. „A Framework for Global Cooperation on Artificial Intelligence and its Governance“. In: *Reflections on Artificial Intelligence for Humanity* (2021), S. 237–265. DOI: 10.1007/978-3-030-69128-8.
- [Ash09] Kevin Ashton. „That ‘Internet of Things’ Thing“. In: *RFID Journal* 22.7 (2009), S. 97–114.
- [Ash76] Robert H. Ashton. „The Robustness of Linear Models for Decision-making“. In: *Omega* 4.5 (1976), S. 609–615. DOI: 10.1016/0305-0483(76)90011-6.
- [AT12] Ivan Antic und Teufik I. Tokic. „RFID: Past, Present, Future“. In: *Scientific Publications of the State University of Novi Pazar* 4.1 (2012), S. 39–52.
- [Ave+20] Pedro H. C. Avelar, Anderson R. Tavares, Thiago L. T. da Silveira, Cláudio R. Jung und Luís C. Lamb. „Superpixel Image Classification with Graph Attention Networks“. In: *33rd IEEE Conference on Graphics, Patterns and Images*. 2020, S. 203–209. DOI: 10.1109/SIBGRAP151738.2020.00035.
- [Bar77] Bruce G. Barrett. „A Further Digression on the over-automated Warehouse: Some Evidence“. In: *Interfaces* 8.1 (1977), S. 46–49. DOI: 10.1287/inte.8.1.46.
- [Bec+21] Katharina Beckh, Sebastian Müller, Matthias Jakobs, Vanessa Toborek, Hanxiao Tan, Raphael Fischer, Pascal Welke, Sebastian Houben und Laura von Rueden. „Explainable Machine Learning with Prior Knowledge: An Overview“. In: *arXiv preprint arXiv:2105.10172* (2021).
- [BH07] Hans-Jörg Bullinger und Michael ten Hompel. *Internet der Dinge: www.internet-der-dinge.de*. Springer, 2007.
- [BH21] Nadia Burkart und Marco F. Huber. „A Survey on the Explainability of Supervised Machine Learning“. In: *Journal of Artificial Intelligence Research* 70 (2021), S. 245–317. DOI: 10.1613/jair.1.12228.
- [Bla79] W. Blanding. „Potential Warehouse Applications of Industrial Robots“. In: *Industrial Robot: An International Journal* 6.3 (1979), S. 125–129. DOI: 10.1108/eb004757.
- [Bor61] Edwin G. Boring. „Intelligence as the Tests Test It“. In: *Studies in Individual Differences: The Search for Intelligence* (1961), S. 210–214. DOI: 10.1037/11491-017.

- [Bro+20] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry und Amanda Askell. „Language Models are Few-shot Learners“. In: *Advances in Neural Information Processing Systems* 33 (2020), S. 1877–1901. DOI: 10.5555/3495724.3495883.
- [Bro99] Thompson Brockman. „21 Warehousing Trends in the 21st Century“. In: *IIE Solutions* 31.7 (1999), S. 36–41.
- [BS22] Patrick Becker und Maximilian Schellert. „eCMR-Update für den Frachtbrief“. In: *Fraunhofer Publica* (2022). DOI: 10.24406/publica-719.
- [Car18] Lara Carminati. „Generalizability in Qualitative Research: A Tale of Two Traditions“. In: *Qualitative Health Research* 28.13 (2018), S. 2094–2101. DOI: 10.1177/1049732318788379.
- [CBB22] Giorgia Casella, Barbara Bigliardi und Eleonora Bottani. „The Evolution of RFID Technology in the Logistics Field: A Review“. In: *Procedia Computer Science* 200 (2022), S. 1582–1592. DOI: 10.1016/j.procs.2022.01.362..
- [CBÖ18] Malgorzata Ciesielska, Katarzyna W. Boström und Magnus Öhlander. „Observation Methods“. In: *Qualitative Methodologies in Organization Studies: Volume II: Methods and Possibilities* (2018), S. 33–52. DOI: 10.1007/978-3-319-65442-3.
- [Čer+24] Vojtěch Čermák, Lukas Pícek, Lukáš Adam und Kostas Papafitsoros. „Wildlife Datasets: An Open-source Toolkit for Animal Re-identification“. In: *IEEE Winter Conference on Applications of Computer Vision*. 2024, S. 5953–5963. DOI: 10.1109/WACV57701.2024.00585.
- [CH24] Simon Caton und Christian Haas. „Fairness in Machine Learning: A Survey“. In: *ACM Computing Surveys* 56.7 (2024), S. 1–38. DOI: 10.1145/3616865.
- [Che+18a] Dapeng Chen, Dan Xu, Hongsheng Li, Nicu Sebe und Xiaogang Wang. „Group Consistent Similarity Learning via Deep CRF for Person Re-identification“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2018, S. 8649–8658. DOI: 10.1109/CVPR.2018.00902.
- [Che+18b] Xuhui Chen, Jinlong Ji, Changqing Luo, Weixian Liao und Pan Li. „When Machine Learning Meets Blockchain: A Decentralized, Privacy-preserving and Secure Design“. In: *IEEE International Conference on Big Data*. 2018, S. 1178–1187. DOI: 10.1109/BigData.2018.8622598.
- [CKS19] Larissa Chazette, Oliver Karras und Kurt Schneider. „Do End-users Want Explanations? Analyzing the Role of Explainability as an Emerging Aspect of Non-functional Requirements“. In: *27th IEEE International Requirements Engineering Conference*. 2019, S. 223–233. DOI: 10.1109/RE.2019.00032.
- [Cla03] Jocalyn P. Clark. „How to Peer Review a Qualitative Manuscript“. In: *Peer Review in Health Sciences* (2003), S. 219–235.

- [CM20] Larissa Custodio und Ricardo Machado. „Flexible Automated Warehouse: A Literature Review and an Innovative Framework“. In: *The International Journal of Advanced Manufacturing Technology* 106 (2020), S. 533–558. DOI: 10.1007/s00170-019-04588-z.
- [CP19] Karolina Czerwińska und Andrzej Pacana. „Analysis of the Implementation of the Identification System for Directly Marked Parts - DataMatrix Code“. In: *Production Engineering Archives* 23.23 (2019), S. 22–26. DOI: 10.30657/pea.2019.23.04.
- [Dal77] Tore Dalenius. „Towards a Methodology for Statistical Disclosure Control“. In: *Statistics Sweden* (1977). DOI: 1813/111303.
- [DCH21] Mohammad Zalbagi Darestani, Akshay S. Chaudhari und Reinhard Heckel. „Measuring Robustness in Deep Learning based Compressive Sensing“. In: *International Conference on Machine Learning*. 2021, S. 2433–2444.
- [De 20] Emiliano De Cristofaro. „An Overview of Privacy in Machine Learning“. In: *arXiv preprint arXiv:2005.08679* (2020).
- [Den+09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li und Li Fei-Fei. „Imagenet: A Large-scale Hierarchical Image Database“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2009, S. 248–255. DOI: 10.1109/CVPR.2009.5206848.
- [DF21] Alessandro Delfanti und Bronwyn Frey. „Humanly Extended Automation or the Future of Work seen through Amazon Patents“. In: *Science, Technology, & Human Values* 46.3 (2021), S. 655–682. DOI: 10.1177/0162243920943665.
- [DIN 13199-1:2000-1] Deutsches Institut für Normung e. V. (Hrsg.). *Verpackung - Kleinladungsträgersysteme - Teil 1: Allgemeine Anforderungen und Prüfverfahren*. Beuth-Verlag, Berlin. 2000.
- [DIN 13698-1:2003-1] Deutsches Institut für Normung e. V. (Hrsg.). *Produktspezifikation für Paletten - Teil 1: Herstellung von 800 mm×1200 mm-Flachpaletten aus Holz*. Beuth-Verlag, Berlin. 2003.
- [DIN 15155:1986-12] Deutsches Institut für Normung e. V. (Hrsg.). *Paletten; Gitterboxpalette mit 2 Vorderwandklappen*. Beuth-Verlag, Berlin. 1986.
- [DIN 9000:2015-11] Deutsches Institut für Normung e. V. (Hrsg.). *DIN 9000:2015-11 - Qualitätsmanagementsysteme; Grundlagen und Begriffe*. Beuth-Verlag, Berlin. 2015.
- [DK05] Martin Dodge und Rob Kitchin. „Codes of Life: Identification Codes and the Machine-readable World“. In: *Environment and Planning D: Society and Space* 23.6 (2005), S. 851–881. DOI: 10.1068/d378t.
- [Dod+19] Jonathan Dodge, Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy und Casey Dugan. „Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment“. In: *24th International Conference on Intelligent User Interfaces*. 2019, S. 275–285. DOI: 10.1145/3301275.3302310.

- [Dre+21] Nathan Drenkow, Numair Sani, Ilya Shpitser und Mathias Unberath. „A Systematic Review of Robustness in Deep Learning for Computer Vision: Mind the Gap?“ In: *arXiv preprint arXiv:2112.00639* (2021).
- [DW16] Pavel Dmitriev und Xian Wu. „Measuring Metrics“. In: *25th ACM International on Conference on Information and Knowledge Management*. 2016, S. 429–437. DOI: 10.1145/2983323.2983356.
- [EBW16] Behzad Esmaeilian, Sara Behdad und Ben Wang. „The Evolution and Future of Manufacturing: A Review“. In: *Journal of Manufacturing Systems* 39 (2016), S. 79–100. DOI: 10.1016/j.jmsy.2016.03.001.
- [ELV07] Andreas Ess, Bastian Leibe und Luc Van Gool. „Depth and Appearance for Mobile Scene Analysis“. In: *11th IEEE International Conference on Computer Vision*. 2007, S. 1–8. DOI: 10.1109/ICCV.2007.4409092.
- [Emm+16] Katharina Emmerich, Natalya Bogacheva, Mareike Bockholt und Viktor Wendel. „Operationalization and Measurement of Evaluation Constructs“. In: *Entertainment Computing and Serious Games: International GI-Dagstuhl Seminar*. Springer. 2016, S. 306–331. DOI: 10.1007/978-3-319-46152-6\_13.
- [EPAL 11/2021] European Pallet Association. *EPAL 11/2021 - EPAL QUALITY CLASSIFICATION; Recommendations for the Use and Exchange of EPAL Euro Pallets*.
- [Erh+10] Dumitru Erhan, Aaron Courville, Yoshua Bengio und Pascal Vincent. „Why Does Unsupervised Pre-training Help Deep Learning?“ In: *13th International Conference on Artificial Intelligence and Statistics*. 2010, S. 201–208. DOI: 10.5555/1756006.1756025.
- [Eur16] European Commission. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*. 2016. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- [Eur23] European Commission. *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 — C9-0146/2021 — 2021/0106(COD))*. 2023. URL: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C\\_202400506](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C_202400506).
- [FAT11] Sergi Foix, Guillem Alenya und Carme Torras. „Lock-in Time-of-flight (ToF) Cameras: A Survey“. In: *IEEE Sensors* 11.9 (2011), S. 1917–1926. DOI: 10.1109/JSEN.2010.2101060.

- [Fer+19] José Fernandes, Francisco J. G. Silva, Raul Campilho, Gustavo F. Pinto und Andresa Baptista. „Intralogistics and Industry 4.0: Designing a Novel Shuttle with Picking System“. In: *Procedia Manufacturing* 38 (2019), S. 1801–1832. DOI: 10.1016/j.promfg.2020.01.078.
- [Fer04] Timothy L. J. Ferris. „A New Definition of Measurement“. In: *Measurement* 36.1 (2004), S. 101–109. DOI: 10.1016/j.measurement.2004.03.001.
- [Fey49] Richard P. Feynman. „Space-Time Approach to Quantum Electrodynamics“. In: *Physical Review* 76 (6 1949), S. 769–789. DOI: 10.1103/PhysRev.76.769.
- [Fin82] Ludwik Finkelstein. *Theory and Philosophy of Measurement*. Bd. 1. John Wiley & Sons, Inc., 1982, S. 1–30. ISBN: 0471100374.
- [FMD19] Chen Fu, Christoph Mertz und John M. Dolan. „Lidar and Monocular Camera Fusion: On-road Depth Completion for Autonomous Driving“. In: *IEEE Intelligent Transportation Systems Conference*. 2019, S. 273–278. DOI: 10.1109/ITSC.2019.8917201.
- [For+21] Nikolaus Forgó, Stefanie Hänold, Jeroen van den Hoven, Tina Krügel, Iryna Lishchuk, René Mahieu, Anna Monreale, Dino Pedreschi, Francesca Pratesi und David van Putten. „An Ethico-legal Framework for Social Data Science“. In: *International Journal of Data Science and Analytics* 11 (2021), S. 377–390. DOI: 10.1007/s41060-020-00211-7.
- [For+99] David A. Forsyth, Joseph L. Mundy, Vito di Gesù, Roberto Cipolla, Yann LeCun, Patrick Haffner, Léon Bottou und Yoshua Bengio. „Object Recognition with Gradient-based Learning“. In: *Shape, Contour and Grouping in Computer Vision* (1999), S. 319–345. DOI: 10.1007/3-540-46805-6\_19.
- [Fu+21] Dengpan Fu, Dongdong Chen, Jianmin Bao, Hao Yang, Lu Yuan, Lei Zhang, Houqiang Li und Dong Chen. „Unsupervised Pre-training for Person Re-identification“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2021, S. 14750–14759. DOI: 10.1109/CVPR46437.2021.01451.
- [Fur+19] Joshua S. Furtado, Hugh H. Liu, Gilbert Lai, Herve Lacheray und Jason Desouza-Coelho. „Comparative Analysis of Optitrack Motion Capture Systems“. In: *Advances in Motion Sensing and Control for Robotic Applications: Selected Papers from the Symposium on Mechatronics, Robotics, and Control*. Springer. 2019, S. 15–31. DOI: 10.1007/978-3-030-17369-2\_2.
- [Gal+20] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt und Kostas Daniilidis. „Event-based Vision: A Survey“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.1 (2020), S. 154–180. DOI: 10.1109/TPAMI.2020.3008413.

- [GBT07] Douglas Gray, Shane Brennan und Hai Tao. „Evaluating Appearance Models for Recognition, Reacquisition, and Tracking“. In: *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*. Bd. 3. 5. 2007.
- [GE07] Russell E. Glasgow und Karen M. Emmons. „How Can We Increase Translation of Research into Practice? Types of Evidence Needed“. In: *Annual Review on Public Health* 28 (2007), S. 413–433. DOI: 10.1146/annurev.publhealth.28.021406.144145.
- [Gei+13] Andreas Geiger, Philip Lenz, Christoph Stiller und Raquel Urtasun. „Vision Meets Robotics: The Kitti Dataset“. In: *The International Journal of Robotics Research* 32.11 (2013), S. 1231–1237. DOI: 10.1177/0278364913491297.
- [Gen11] Jason Geng. „Structured-light 3D Surface Imaging: A Tutorial“. In: *Advances in Optics and Photonics* 3.2 (2011), S. 128–160. DOI: 10.1364/AOP.3.000128.
- [GF20] Chetan Gupta und Ahmed Farahat. „Deep Learning for Industrial AI: Challenges, New Methods and Best Practices“. In: *26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, S. 3571–3572. DOI: 10.1145/3394486.3406482.
- [GH10] Willibald Günthner und Michael ten Hompel. *Internet der Dinge in der Intralogistik*. Springer, 2010. DOI: 10.1007/978-3-642-04896-8.
- [GMP18] Ian Goodfellow, Patrick McDaniel und Nicolas Papernot. „Making Machine Learning Robust Against Adversarial Inputs“. In: *Communications of the ACM* 61.7 (2018), S. 56–66. DOI: 10.1145/3134599.
- [Goo+20] Dou Goodman, Hao Xin, Wang Yang, Wu Yuesheng, Xiong Junfeng und Zhang Huan. „Advbox: A Toolbox to Generate Adversarial Examples that Fool Neural Networks“. In: *arXiv preprint arXiv:2001.05574* (2020).
- [Gou+21] Anas Gouda, Abraham Ghanem, Pascal Kaiser und Michael ten Hompel. „Object Class-agnostic Segmentation for Practical CNN Utilization in Industry“. In: *6th International Conference on Mechanical Engineering and Robotics Research*. 2021, S. 97–105. DOI: 10.1109/ICMERR54363.2021.9680821.
- [Gou13] Stephen Goundrey-Smith. „Barcodes and Logistics“. In: *Information Technology in Pharmacy: An Integrated Approach* (2013), S. 175–192. DOI: 10.1007/978-1-4471-2780-2\_7.
- [GP08] Fosca Giannotti und Dino Pedreschi. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Springer Science & Business Media, 2008. DOI: 10.1007/978-3-540-75177-9.
- [GP17] Ian Goodfellow und Nicolas Papernot. „The Challenge of Verification and Testing of Machine Learning“. In: *Cleverhans-blog* (2017). URL: <http://www.cleverhans.io/security/privacy/ml/2017/06/14/verification.html>.
- [GS1 DMG] GS1 AISBL. *GS1 DataMatrix Guideline Overview and Technical Introduction to the Use of GS1 DataMatrix*.

- [GS23] Alexander Goudz und Henrik Sanchez-Gonzalez. „Das semantische Internet: Die Brücke zwischen physischen und digitalen Gütern“. In: *Towards the New Normal in Mobility: Technische und betriebswirtschaftliche Aspekte*. Springer, 2023, S. 595–613. DOI: 10.1007/978-3-658-39438-7\_34.
- [Gud04] Timm Gudehus. *Logistik: Grundlagen - Strategien - Anwendungen*. Springer, 2004. DOI: 10.1007/978-3-662-08408-3.
- [Gud05] Timm Gudehus. *Logistik: Grundlagen - Strategien - Anwendungen*. Springer, 2005. DOI: 10.1007/3-540-27629-7.
- [Gui+18] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti und Dino Pedreschi. „A Survey of Methods for Explaining Black Box Models“. In: *ACM Computing Surveys* 51.5 (2018), S. 1–42. DOI: 10.1145/3236009.
- [GW97] Werner Greve und Dirk Wentura. *Wissenschaftliche Beobachtung: Eine Einführung*. Beltz Verlag, 1997. ISBN: 3621273603.
- [Har06] Cyril M. Harris. *Dictionary of Architecture & Construction*. McGraw-Hill Education Ltd., 2006. ISBN: 0071452370.
- [Har19] Masahiro Hara. „Development and Popularization of the QR-Code — Code Development Pursuing Reading Performance and Market Forming by Open Strategy“. In: *Synthesiology (English Edition)* 12.1 (2019), S. 19–28. DOI: 10.5571/syntheng.12.1\_19.
- [Har93] Roy L. Harmon. *Reinventing the Warehouse: World Class Distribution Logistics*. Simon & Schuster, 1993. ISBN: 0029138639.
- [HBL17] Alexander Hermans, Lucas Beyer und Bastian Leibe. „In Defense of the Triplet Loss for Person Re-identification“. In: *arXiv preprint arXiv:1703.07737* (2017).
- [HC88] John R. Hauser und Don Clausing. *The House of Quality*. Bd. 66. Harvard Business School, 1988. ISBN: 0000883077.
- [He+16] Kaiming He, Xiangyu Zhang, Shaoqing Ren und Jian Sun. „Deep Residual Learning for Image Recognition“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016, S. 770–778. DOI: 10.1109/CVPR.2016.90.
- [HE22] Abhinav Hasija und Terry L. Esper. „In Artificial Intelligence (AI) We Trust: A Qualitative Investigation of AI Technology Acceptance“. In: *Journal of Business Logistics* 43.3 (2022), S. 388–412. DOI: 10.1111/jbl.12301.
- [Hee+20] Jessica Heesen, Jörn Müller-Quade, Stefan Wrobel, Maximilian Poretschkin, Stephanie Dachsberger und Maximilian Hösl. „Zertifizierung von KI-Systemen – Impulspapier“. In: *Plattform Lernende Systeme* (2020).

- [Hen+14] Jean-Marc Henin, Michael Leyman, Aline Bauduin, Benoit Jourez und Jacques Hébert. „Phyosanitary Treatment of European Pallets by Microwave: Developing a Program to Ensure Compliance with ISPM 15 and Monitoring its Efficacy on the House Longhorn Beetle“. In: *European Journal of Wood and Wood Products* 72.5 (2014), S. 623–633. DOI: 10.1007/s00107-014-0826-4.
- [Her+21] Lukas-Valentin Herm, Jonas Wanner, Franz Seubert und Christian Janiesch. „I Don’t Get it, but it Seems Valid! The Connection between Explainability and Comprehensibility in (X)AI Research“. In: *European Conference on Information Systems*. 2021.
- [Her+23] Lukas-Valentin Herm, Kai Heinrich, Jonas Wanner und Christian Janiesch. „Stop Ordering Machine Learning Algorithms by Their Explainability! A User-centered Investigation of Performance and Explainability“. In: *International Journal of Information Management* 69 (2023), S. 102538. DOI: 10.1016/j.ijinfomgt.2022.102538.
- [HH14] Michael ten Hompel und Michael Henke. „Logistik 4.0“. In: *Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung - Technologien - Migration* (2014), S. 615–624. DOI: 10.1007/978-3-658-04682-8.
- [HH22] Azmi Haider und Hagit Hel-Or. „What Can We Learn From Depth Camera Sensor Noise?“ In: *Sensors* 22.14 (2022), S. 5448. DOI: 10.3390/s22145448.
- [Hol+20] Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek und Wojciech Samek. „Explainable AI Methods – A Brief Overview“. In: *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Springer. 2020, S. 13–38. DOI: 10.1007/978-3-031-04083-2\_2.
- [Hol+23] Dimitrij-Marian Holm, Philipp Junge, Jérôme Rutinowski und Johannes Fottner. „Investigation of Deep Learning Datasets for Intralogistics“. In: *Conference on Production Systems and Logistics*. 2023. DOI: 10.15488/15311.
- [Hom+20] Michael ten Hompel, Hacı Bayhan, Jan Behling, Leonie Benkenstein, Jan Emmerich, Guido Follert, Martin Grzenia, Christian Hammermeister, Hendrik Hasse und Dirk Hoening. „Technical Report: LoadRunner®, a New Platform Approach on Collaborative Logistics Services“. In: *Logistics Journal* 2020.10 (2020). DOI: 10.2195/lj\_NotRev\_tenhompel\_en\_202010\_01.
- [Hom06] Michael ten Hompel. „Das Internet der Dinge - Potenziale autonomer Objekte und selbstorganisierender Systeme in der Intralogistik“. In: *Intralogistik - Potentiale, Perspektiven, Prognosen* (2006), S. 266–276. DOI: 10.1007/978-3-540-29658-4.
- [Hon09] Sun Hong-Ying. „The Application of Barcode Technology in Logistics and Warehouse Management“. In: *1st IEEE International Workshop on Education Technology and Computer Science*. Bd. 3. 2009, S. 732–735. DOI: 10.1109/ETCS.2009.698.

- [Hou+22a] Ruitao Hou, Shan Ai, Qi Chen, Hongyang Yan, Teng Huang und Kongyang Chen. „Similarity-based Integrity Protection for Deep Learning Systems“. In: *Information Sciences* 601 (2022), S. 255–267. DOI: 10.1016/j.ins.2022.04.003.
- [Hou+22b] Sebastian Houben, Stephanie Abrecht, Maram Akila, Andreas Bär, Felix Brockherde, Patrick Feifel, Tim Fingscheidt, Sujan S. Gannamaneni, Seyed E. Ghobadi und Ahmed Hammam. „Inspect, Understand, Overcome: A Survey of Practical Methods for AI Safety“. In: *Deep Neural Networks and Data for Automated Driving: Robustness, Uncertainty Quantification, and Insights Towards Safety*. Springer, 2022, S. 3–78. DOI: 10.1007/978-3-031-01233-4\_1.
- [HPM18] Sibylle Hess, Nico Piatkowski und Katharina Morik. „The Trustworthy Pal: Controlling the False Discovery Rate in Boolean Matrix Factorization“. In: *SIAM International Conference on Data Mining*. 2018, S. 405–413. DOI: 10.1137/1.9781611975321.46.
- [HPS16] Moritz Hardt, Eric Price und Nathan Srebro. „Equality of Opportunity in Supervised Learning“. In: *30th International Conference on Neural Information Processing Systems*. 2016. DOI: 10.5555/3157382.3157469.
- [HR97] Timothy Huang und Stuart Russell. „Object Identification in a Bayesian Context“. In: *International Joint Conferences on Artificial Intelligence*. Bd. 97. 1997, S. 1276–1282. DOI: 10.5555/1622270.1622340.
- [HS91] Stuart Haber und Scott Stornetta. „How to Time-stamp a Digital Document“. In: *Journal of Cryptology* 3.2 (1991), S. 99–111. DOI: 10.1007/BF00196791.
- [Hua+20] Xiaowei Huang, Daniel Kroening, Wenjie Ruan, James Sharp, Youcheng Sun, Emese Thamo, Min Wu und Xinpeng Yi. „A Survey of Safety and Trustworthiness of Deep Neural Networks: Verification, Testing, Adversarial Attack and Defence, and Interpretability“. In: *Computer Science Review* 37 (2020). DOI: 10.1016/j.cosrev.2020.100270.
- [Hut+21] Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes und Margaret Mitchell. „Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure“. In: *ACM Conference on Fairness, Accountability, and Transparency*. 2021, S. 560–575. DOI: 10.1145/3442188.3445918.
- [Hut18] Matthew Hutson. „Artificial Intelligence Faces Reproducibility Crisis“. In: *American Association for the Advancement of Science* (2018). DOI: 10.1126/science.359.6377.725.
- [HW11] John C. Hermanson und Alex C. Wiedenhoeft. „A Brief Review of Machine Vision in the Context of Automated Wood Identification Systems“. In: *IAWA Journal* 32.2 (2011), S. 233–250. DOI: 10.1163/22941932-90000054.

- [HZL18] Zecheng He, Tianwei Zhang und Ruby B. Lee. „Verideep: Verifying Integrity of Deep Neural Networks through Sensitive-sample Fingerprinting“. In: *arXiv preprint arXiv:1808.03277* (2018).
- [ISO/IEC 15420:2009] International Organization for Standardization (ISO). *ISO/IEC 15420:2009 Information Technology Automatic Identification and Data Capture Techniques EAN/UPC Bar Code Symbology Specification*.
- [Ist+20] Nadya A. Istiqomah, Putri F. Sansabilla, Doddy Himawan und Muhammad Rifni. „The Implementation of Barcode on Warehouse Management System for Warehouse Efficiency“. In: *Journal of Physics: Conference Series* 1573.1 (2020). DOI: 10.1088/1742-6596/1573/1/012038.
- [Jav+21] Mohd Javaid, Abid Haleem, Shanay Rab, Ravi P. Singh und Rajiv Suman. „Sensors for Daily Life: A Review“. In: *Sensors International* 2 (2021). DOI: 10.1016/j.sintl.2021.100121.
- [Jew+16] Zoe Jewell, Sky Alibhai, Florian Weise, Stuart Munro, Marlice Vuuren und Rudie Vuuren. „Spotting Cheetahs: Identifying Individuals by their Footprints“. In: *Journal of Visualized Experiments* (2016). DOI: 10.3791/54034.
- [JLE14] Zhanglong Ji, Zachary C. Lipton und Charles Elkan. „Differential Privacy and Machine Learning: A Survey and Review“. In: *arXiv preprint arXiv:1412.7584* (2014).
- [JSS17] Thorsten Joachims, Adith Swaminathan und Tobias Schnabel. „Unbiased Learning-to-rank with Biased Feedback“. In: *10th ACM International Conference on Web Search and Data Mining*. 2017, S. 781–789. DOI: 10.1145/3018661.3018699.
- [Kam+11] Andreas Kamagaew, Jonas Stenzel, Andreas Nettsträter und Michael ten Hompel. „Concept of Cellular Transport Systems in Facility Logistics“. In: *The 5th International Conference on Automation, Robotics and Applications*. 2011, S. 40–45. DOI: 10.1109/ICARA.2011.6144853.
- [KK99] Akio Kosaka und Avinash C. Kak. *Stereo Vision for Industrial Applications*. John Wiley & Sons, Inc., 1999. ISBN: 0471177830.
- [Kla95] Anton Klar. „Identifizierung von Ladehilfsmitteln auf der Basis Stochastischer Oberflächeninformationen“. In: *Verlag Praxiswissen*. 1995. ISBN: 3-929443-53-8.
- [Klü+22] Simon Klüttermann, Jérôme Rutinowski, Christopher Reining, Moritz Roidl und Emmanuel Müller. „Towards Graph Representation Based Re-identification of Chipwood Pallet Blocks“. In: *21st IEEE International Conference on Machine Learning and Applications*. 2022, S. 1543–1550. DOI: 10.1109/ICMLA55696.2022.00279.

- [Klü+23] Simon Klüttermann, Jérôme Rutinowski, Anh Nguyen, Christopher Reining, Moritz Roidl und Emmanuel Müller. „On Graph Representation based Re-Identification – A Proof of Concept“. In: *IEEE International Conference on Data Mining Workshops (ICDMW)*. 2023, S. 1097–1104. DOI: 10.1109/ICDMW60847.2023.00144.
- [Klü+24] Simon Klüttermann, Jérôme Rutinowski, Anh Nguyen, Britta Grimme, Moritz Roidl und Emmanuel Müller. „On the Effectiveness of Heterogeneous Ensemble Methods for Re-identification“. In: *arXiv preprint arXiv:2403.12606* (2024).
- [KLZ05] Ekart Kuhn, Volker Lange und Peter Zimmermann. *Paletten-Management: Leitfaden für die Praxis*. Verlag Heinrich Vogel, 2005. ISBN: 3574260466.
- [KMS22] Khadija Khaldi, Pranav Mantini und Shishir K. Shah. „Un-supervised Person Re-identification Based on Skeleton Joints Using Graph Convolutional Networks“. In: *Image Analysis and Processing*. Springer, 2022, S. 135–146. DOI: 10.1007/978-3-031-06433-3\_12.
- [Kon+19] Linghuan Kong, Wei He, Chenguang Yang, Zhijun Li und Changyin Sun. „Adaptive Fuzzy Control for Coordinated Multiple Robots with Constraint using Impedance Learning“. In: *IEEE Transactions on Cybernetics* 49.8 (2019), S. 3052–3063. DOI: 10.1109/TCYB.2018.2838573.
- [Kor+20] Yassin Kortli, Maher Jridi, Ayman Al Falou und Mohamed Atri. „Face Recognition Systems: A Survey“. In: *Sensors* 20.2 (2020). DOI: 10.3390/s20020342.
- [KP20] Ladislav Karrach und Elena Pivarčiová. „Recognition of Data Matrix Codes in Images and their Applications in Production Processes“. In: *Management Systems in Production Engineering* 28.3 (2020), S. 154–161. DOI: 10.2478/mspe-2020-0023.
- [KR18] Eline van der Kruk und Marco M. Reijne. „Accuracy of Human Motion Capture Systems for Sport Applications; State-of-the-art Review“. In: *European Journal of Sport Science* 18.6 (2018), S. 806–819. DOI: 10.1080/17461391.2018.1463397.
- [Krä+22] Larissa Krämer, Jérôme Rutinowski, Jan Endendyk, Robert Vaut und Moritz Roidl. „On Blockchain-based Token Usage in Cyber-Physical Production Systems“. In: *Logistics Journal: Proceedings* (2022). DOI: 10.2195/lj\_proc\_kraemer\_en\_202211\_01.
- [KRM24] Simon Klüttermann, Jérôme Rutinowski und Emmanuel Müller. „The Phenomenon of Correlated Representations in Contrastive Learning“. In: *IEEE International Joint Conference on Neural Networks*. 2024. DOI: 10.1109/IJCNN60899.2024.10649913.

- [Kro+14] Katharina Krombholz, Peter Frühwirt, Peter Kieseberg, Ioannis Kapsalis, Markus Huber und Edgar Weippl. „QR Code Security: A Survey of Attacks and Challenges for Usable Security“. In: *2nd International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer. 2014, S. 79–90. DOI: 10.1007/978-3-319-07620-1\_8.
- [Kus+17] Matt J. Kusner, Joshua Loftus, Chris Russell und Ricardo Silva. „Counterfactual Fairness“. In: *Advances in Neural Information Processing Systems*. Bd. 30. Curran Associates, Inc., 2017. DOI: 10.5555/3294996.3295162.
- [Kut+22] Deepthi P. Kuttichira, Sunil Gupta, Dang Nguyen, Santu Rana und Svetha Venkatesh. „Verification of Integrity of Deployed Deep Learning Models using Bayesian Optimization“. In: *Knowledge-based Systems* 241 (2022). DOI: 10.1016/j.knosys.2022.108238.
- [Lat15] Wallace Latimer. „Understanding Laser-based 3D Triangulation Methods“. In: *Vision Systems Design* 20.6 (2015).
- [LeC+98] Yann LeCun, Léon Bottou, Yoshua Bengio und Patrick Haffner. „Gradient-based Learning Applied to Document Recognition“. In: *Proceedings of the IEEE* 86.11 (1998), S. 2278–2324. DOI: 10.1109/5.726791.
- [Lei86] Gottfried Wilhelm Leibniz. *Discours de métaphysique*. 1686.
- [Lep+18] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland und Patrick Vinck. „Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges“. In: *Philosophy & Technology* 31 (2018), S. 611–627. DOI: 10.1007/s13347-017-0279-x.
- [Lev+21] Elad Levi, Tete Xiao, Xiaolong Wang und Trevor Darrell. „Rethinking Preventing Class-collapsing in Metric Learning with Margin-based Losses“. In: *IEEE International Conference on Computer Vision*. 2021, S. 10316–10325. DOI: 10.1109/ICCV48922.2021.01015.
- [LHG12] Ryan Layne, Timothy M. Hospedales und Shaogang Gong. „Person Re-identification by Attributes.“ In: *Queen Mary Vision Laboratory, School of Electronic Engineering and Computer Science*. 2012. DOI: 10.5244/C.26.24.
- [Li+14] Wei Li, Rui Zhao, Tong Xiao und Xiaogang Wang. „DeepReID: Deep Filter Pairing Neural Network for Person Re-identification“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2014, S. 152–159. DOI: 10.1109/CVPR.2014.27.
- [Li+18] Binbin V. Li, Sky Alibhai, Zoe Jewell, Desheng Li und Hemin Zhang. „Using Footprints to Identify and Sex Giant Pandas“. In: *Biological Conservation* 218 (2018), S. 83–90. DOI: 10.1016/j.biocon.2017.11.029.

- [Li+19] Tianjian Li, Qiu Jin, Bin Huang, Chang Li und Min Huang. „Cargo Pallets Real-time 3D Positioning Method Based on Computer Vision“. In: *7th International Symposium on Test Automation and Instrumentation*. 2019. DOI: 10.1049/joe.2018.9053.
- [Li+21] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng und Jun Zhou. „A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects“. In: *IEEE Transactions on Neural Networks and Learning Systems* (2021). DOI: 10.1109/TNNLS.2021.3084827.
- [Li+23] Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi und Bowen Zhou. „Trustworthy AI: From Principles to Practices“. In: *ACM Computing Surveys* 55.9 (2023), S. 1–46. DOI: 10.1145/3555803.
- [Lie15] Thomas Liebig. „Privacy Preserving Centralized Counting of Moving Objects“. In: *Geographic Information Science as an Enabler of Smarter Cities and Communities* (2015), S. 91–103. DOI: 10.1007/978-3-319-16787-9\_6.
- [Lin+14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár und Lawrence Zitnick. „Microsoft COCO: Common Objects in Context“. In: *13th European Conference on Computer Vision*. Springer. 2014, S. 740–755. DOI: 10.1007/978-3-319-10602-1\_48.
- [Liu+15] Ziwei Liu, Ping Luo, Xiaogang Wang und Xiaoou Tang. „Deep Learning Face Attributes in the Wild“. In: *IEEE International Conference on Computer Vision*. 2015, S. 3730–3738. DOI: 10.1109/ICCV.2015.425.
- [Liu+20] Yiming Liu, Richard Yu, Xi Li, Hong Ji und Victor Leung. „Blockchain and Machine Learning for Communications and Networking Systems“. In: *IEEE Communications Surveys & Tutorials* 22.2 (2020), S. 1392–1431. DOI: 10.1109/COMST.2020.2975911.
- [Liu+23] Jiawan Liu, Duojin Wang, Qinglian Lin und Meikui Deng. „Risk Assessment based on FMEA Combining DEA and Cloud Models: A Case Application in Robot-assisted Rehabilitation“. In: *Expert Systems with Applications* 214 (2023). DOI: 10.1016/j.eswa.2022.119119.
- [LLL13] Hu-Chen Liu, Long Liu und Nan Liu. „Risk Evaluation Approaches in Failure Mode and Effects Analysis: A Literature Review“. In: *Expert Systems with Applications* 40.2 (2013), S. 828–838. DOI: 10.1016/j.eswa.2012.08.010.
- [LLP12] James Ladyman, Øystein Linnebo und Richard Pettigrew. „Identity and Discernibility in Philosophy and Logic“. In: *The Review of Symbolic Logic* 5.1 (2012), S. 162–186. DOI: 10.1017/S1755020311000281.

- [Lok+14] Namita Lokare, Qian Ge, Wesley Snyder, Zoe Jewell, Sky Alibhai und Edgar Lobaton. „Manifold Learning Approach to Curve Identification with Applications to Footprint Segmentation“. In: *IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing*. 2014, S. 1–8. DOI: 10.1109/CIMSIVP.2014.7013288.
- [Lor23] Fabian J. Lorson. „Perspectives on the Warehouse of the Future“. In: *Technische Universität München*. 2023.
- [LS16] Amy van Looy und Aygun Shafagatova. „Business Process Performance Measurement: A Structured Literature Review of Indicators, Measures and Metrics“. In: *Springer* 5.1 (2016), S. 1–24. DOI: 10.1186/s40064-016-3498-1.
- [Luo+19] Cai Luo, Leijian Yu, Erfu Yang, Huiyu Zhou und Peng Ren. „A Benchmark Image Dataset for Industrial Tools“. In: *Pattern Recognition Letters* 125 (2019), S. 341–348. DOI: 10.1016/j.patrec.2019.05.011.
- [Mat+23] Stephan Matzke, Christian Pionzewski, Jérôme Rutinowski, Firdes Celik und Markus König. „Re-identification Robustness Over Time – The Case for Synthetic Training Data“. In: *International Conference on Machine Learning and Applications*. 2023, S. 1363–1370. DOI: 10.1109/ICMLA58977.2023.00206.
- [Mat24] Anjali Mathur. *Rising of Dark Factories due to Artificial Intelligence*. CRC Press, 2024, S. 304–314. ISBN: 9781003432319.
- [May+20] Christopher Mayershofer, Dimitrij-Marian Holm, Benjamin Molter und Johannes Fottner. „LOCO: Logistics Objects in Context“. In: *19th IEEE International Conference on Machine Learning and Applications*. 2020, S. 612–617. DOI: 10.1109/ICMLA51294.2020.00102.
- [MB20] Christian Meske und Enrico Bunde. „Transparency and Trust in Human-AI-interaction: The Role of Model-agnostic Explanations in Computer Vision-based Decision Support“. In: *First International Conference on Artificial Intelligence in HCI*. Springer. 2020, S. 54–69. DOI: 10.1007/978-3-030-50334-5\_4.
- [McK83] Alan C. McKinnon. „The Development of Warehousing in England“. In: *Geoforum* 14.4 (1983), S. 389–399. DOI: 10.1016/0016-7185(83)90036-2.
- [MDS95] Roger C. Mayer, James H. Davis und David Schoorman. „An Integrative Model of Organizational Trust“. In: *Academy of Management Review* 20.3 (1995), S. 709–734. DOI: 10.2307/258792.
- [MF18] Benjamin Molter und Johannes Fottner. „Real-time Pallet Localization with 3D Camera Technology for Forklifts in Logistic Environments“. In: *IEEE International Conference on Service Operations and Logistics, and Informatics*. 2018, S. 297–302. DOI: 10.1109/SOLI.2018.8476740.
- [MI15] Pavan Meadati und Javier Irizarry. „BIM and QR code for Operation and Maintenance“. In: *Computing in Civil Engineering* 2015. 2015, S. 556–563. DOI: 10.1061/9780784479247.069.

- [Mic97] Joel Michell. „Quantitative Science and the Definition of Measurement in Psychology“. In: *British journal of Psychology* 88.3 (1997), S. 355–383. DOI: 10.1111/j.2044-8295.1997.tb02641.x.
- [Min+22] Zhangqiang Ming, Min Zhu, Xiangkun Wang, Jiamin Zhu, Junlong Cheng, Chengrui Gao, Yong Yang und Xiaoyong Wei. „Deep Learning-Based Person Re-Identification Methods: A Survey and Outlook of Recent Works“. In: *Image and Vision Computing* 119 (2022). DOI: 10.1016/j.imavis.2022.104394.
- [Mit+18] Anton Mitrokhin, Cornelia Fermüller, Chethan Parameshwara und Yiannis Aloimonos. „Event-based Moving Object Detection and Tracking“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2018, S. 1–9. DOI: 10.1109/IROS.2018.8593805.
- [Mit+19] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa D. Raji und Timnit Gebru. „Model Cards for Model Reporting“. In: *ACM Conference on Fairness, Accountability, and Transparency*. 2019, S. 220–229. DOI: 10.1145/3287560.3287596.
- [Mit80] Tom M. Mitchell. „The Need for Biases in Learning Generalizations“. In: *CMU School of Computer Science* (1980).
- [Moh+22] Sina Mohseni, Haotao Wang, Chaowei Xiao, Zhiding Yu, Zhangyang Wang und Jay Yadawa. „Taxonomy of Machine Learning Safety: A Survey and Primer“. In: *ACM Computing Surveys* 55.8 (2022), S. 1–38. DOI: 10.1145/3551385.
- [Mon+19] Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek und Klaus-Robert Müller. „Layer-wise Relevance Propagation: An Overview“. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (2019), S. 193–209. DOI: 10.1007/978-3-030-28954-6\_10.
- [Mor+21] Katharina Morik, Helena Kotthaus, Lukas Heppe, Danny Heinrich, Raphael Fischer, Andreas Pauly und Nico Piatkowski. „The Care Label Concept: A Certification Suite for Trustworthy and Resource-aware Machine Learning“. In: *arXiv preprint arXiv:2106.00512* (2021).
- [Mor+22] Katharina Morik, Helena Kotthaus, Raphael Fischer, Sascha Mücke, Matthias Jakobs, Nico Piatkowski, Andreas Pauly, Lukas Heppe und Danny Heinrich. „Yes We Care! - Certification for Machine Learning Methods through the Care Label Framework“. In: *Frontiers in Artificial Intelligence* 5 (2022). DOI: 10.3389/frai.2022.975029.
- [MSM18] Grégoire Montavon, Wojciech Samek und Klaus-Robert Müller. „Methods for Interpreting and Understanding Deep Neural Networks“. In: *Digital Signal Processing* 73 (2018), S. 1–15. DOI: 10.1016/j.dsp.2017.10.011.
- [Muc+18] Claus Muchna, Hans Brandenburg, Johannes Fottner und Jens Gutermuth. *Grundlagen der Logistik: Begriffe, Strukturen und Prozesse*. Springer, 2018. DOI: 10.1007/978-3-658-30836-0.

- [MV20] Ričards Marcinkevičs und Julia E. Vogt. „Interpretability and Explainability: A Machine Learning Zoo Mini-tour“. In: *arXiv preprint arXiv:2012.01805* (2020).
- [NBN19] Athira Nambiar, Alexandre Bernardino und Jacinto C. Nascimento. „Gait-Based Person Re-Identification: A Survey“. In: *ACM Computing Surveys* 52.2 (2019), S. 1–34. DOI: 10.1145/3243043.
- [Nie+20] Friedrich Niemann, Christopher Reining, Fernando Moya Rueda, Nilah Ravi Nair, Janine A. Steffens, Gernot A. Fink und Michael ten Hompel. „LARA: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes“. In: *Sensors* 20.15 (2020). DOI: 10.3390/s20154083.
- [Pap+18] Nicolas Papernot, Patrick McDaniel, Arunesh Sinha und Michael P. Wellman. „SoK: Security and Privacy in Machine Learning“. In: *IEEE European Symposium on Security and Privacy*. 2018, S. 399–414. DOI: 10.1109/EuroSP.2018.00035.
- [PF23] Christoph Pott und Felix Feldmann. „Lagerverwaltung und Lagersteuerung“. In: *IT und autonome Systeme in der Logistik*. Springer, 2023, S. 17–28. DOI: 10.1007/978-3-662-66939-6\_2.
- [Pip14] Alexander Piperski. „An Application of Graph Theory to Linguistic Complexity“. In: *The Poznan Linguistic Meeting 1* (Dez. 2014). DOI: 10.1515/yplm-2015-0005.
- [Pla95] Charlie Plain-Jones. „Data Matrix Identification“. In: *Sensor Review* 15.1 (1995), S. 12–15. DOI: 10.1108/EUM0000000004265.
- [PLW02] Leo L. Pipino, Yang W. Lee und Richard Y. Wang. „Data quality assessment“. In: *Communications of the ACM* 45.4 (2002), S. 211–218. DOI: 10.1145/505248.506010.
- [PMN21] Kenneth Peng, Arunesh Mathur und Arvind Narayanan. „Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers“. In: *35th International Conference on Neural Information Processing Systems*. 2021.
- [Pri20] Christian Priebe. „Protecting Applications using Trusted Execution Environments“. In: *Imperial College London*. 2020.
- [PS22] Dana Pessach und Erez Shmueli. „A Review on Fairness in Machine Learning“. In: *ACM Computing Surveys* 55.3 (2022), S. 1–44. DOI: 10.1145/3494672.
- [PSW00] Raja Parasuraman, Thomas B. Sheridan und Christopher D. Wickens. „A Model for Types and Levels of Human Interaction with Automation“. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 30.3 (2000), S. 286–297. DOI: 10.1109/3468.844354.
- [Raf+21] Hasan Rafiq, Xiaohan Shi, Hengxu Zhang, Huimin Li, Manesh K. Ochani und Aamer A. Shah. „Generalizability Improvement of Deep Learning-based Non-intrusive Load Monitoring System using Data Augmentation“. In: *IEEE Transactions on Smart Grid* 12.4 (2021), S. 3265–3277. DOI: 10.1109/TSG.2021.3082622.

- [Raj+20] Inioluwa D. Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron und Parker Barnes. „Closing the AI Accountability Gap: Defining an End-to-end Framework for Internal Algorithmic Auditing“. In: *ACM Conference on Fairness, Accountability, and Transparency*. 2020, S. 33–44. DOI: 10.1145/3351095.3372873.
- [RBB17] Jonas Rauber, Wieland Brendel und Matthias Bethge. „Foolbox: A Python Toolbox to Benchmark the Robustness of Machine Learning Models“. In: *arXiv preprint arXiv:1707.04131* (2017).
- [Red+16] Joseph Redmon, Santosh Divvala, Ross Girshick und Ali Farhadi. „You Only Look Once: Unified, Real-time Object Detection“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016, S. 779–788. DOI: 10.1109/CVPR.2016.91.
- [Rei+23] Christopher Reining, Nilah Ravi Nair, Friedrich Niemann, Fernando Moya Rueda und Gernot A. Fink. „A Tutorial on Dataset Creation for Sensor-based Human Activity Recognition“. In: (2023), S. 453–459. DOI: 10.1109/PerComWorkshops56833.2023.10150401.
- [Ric+20] John Richards, David Piorkowski, Michael Hind, Stephanie Houde und Aleksandra Mojsilović. „A Methodology for Creating AI FactSheets“. In: *arXiv preprint arXiv:2006.13796* (2020).
- [Ric71] Geoffrey Rickman. *Roman Granaries and Store Buildings*. CUP Archive, 1971. ISBN: 0521077249.
- [Ris+16] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara und Carlo Tomasi. „Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking“. In: *European Conference on Computer Vision Workshop on Benchmarking Multi-Target Tracking*. 2016. DOI: 10.1007/978-3-319-48881-3\_2.
- [Roi22] Moritz Roidl. „Zur innerbetrieblichen Logistik – Axiomatik und Betrachtung als kinodynamisches System“. In: *Verlag Praxiswissen*. 2022. DOI: 10.17877/DE290R-23033.
- [Ros16] Antje Rosenthal. „Systematisierung der Ladungsträgerarten“. In: *Ganzheitliche Bewertung modularer Ladungsträgerkonzepte: Eine Lebenszyklusbetrachtung* (2016), S. 11–30. DOI: 10.1007/978-3-658-15676-3\_2.
- [Ros89] Edward E. Roskam. „Operationalization, a Superfluous Concept“. In: *Quality & Quantity* 23 (1989), S. 237–275. DOI: 10.1007/BF00172446.
- [RS19] Committee on Reproducibility und Replicability in Science. „Reproducibility and Replicability in Science“. In: *National Academies Press* (2019). DOI: 10.17226/25303.
- [RSR12] A. Ramaa, K. N. Subramanya und T. M. Rangaswamy. „Impact of Warehouse Management System in a Supply Chain“. In: *International Journal of Computer Applications* 54.1 (2012). DOI: 10.5120/8530-2062.

- [Rue+21] Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick und Rajkumar Ramamurthy. „Informed Machine Learning - A Taxonomy and Survey of Integrating prior Knowledge into Learning Systems“. In: *IEEE Transactions on Knowledge and Data Engineering* 35.1 (2021), S. 614–633. DOI: 10.1109/TKDE.2021.3079836.
- [Rüp06] Stefan Rüping. „Learning Interpretable Models“. In: *Technische Universität Dortmund*. 2006. DOI: 10.17877/DE290R-8863.
- [Rut+21] Jérôme Rutinowski, Christian Pionzewski, Tim Chilla, Christopher Reining und Michael ten Hompel. „Towards Re-Identification for Warehousing Entities - A Work-in-Progress Study“. In: *26th IEEE International Conference on Emerging Technologies and Factory Automation*. 2021, S. 501–504. DOI: 10.1109/ETFA45728.2021.9613250.
- [Rut+22a] Jérôme Rutinowski, Christian Pionzewski, Tim Chilla, Christopher Reining und Michael ten Hompel. „Deep Learning Based Re-identification of Wooden Euro-pallets“. In: *21st IEEE International Conference on Machine Learning and Applications*. 2022, S. 113–117. DOI: 10.1109/ICMLA55696.2022.00023.
- [Rut+22b] Jérôme Rutinowski, Bhargav Vankayalapati, Nils Schwenzfeier, Maribel Acosta und Christopher Reining. „On the Applicability of Synthetic Data for Re-Identification“. In: *AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*. 2022. DOI: arXiv:2212.10105.
- [Rut+22c] Jérôme Rutinowski, Hazem Youssef, Anas Gouda, Christopher Reining und Moritz Roidl. „The Potential of Deep Learning based Computer Vision in Warehousing Logistics“. In: *Logistics Journal: Proceedings* (2022). DOI: 10.2195/lj\_proc\_rutinowski\_en\_202211\_01.
- [Rut+23] Jérôme Rutinowski, David Schuster, Sönke Kauffmann, Pedro Arias-Pérez, Frederik Polachowski, Miguel Granero, Moritz Roidl und Pascual Campoy. „Exploring the Re-Identification of Industrial Entities on Autonomous Guided Vehicles“. In: *Logistics Journal: Referierte Veröffentlichungen* (2023). DOI: 10.2195/lj\_proc\_rutinowski\_en\_202310\_01.
- [Rut+24a] Jérôme Rutinowski, Sven Franke, Jan Endendyk, Ina Dormuth, Moritz Roidl und Markus Pauly. „The Self-perception and Political Biases of ChatGPT“. In: *Human Behavior and Emerging Technologies 2024* (2024). DOI: 10.1155/2024/7115633.
- [Rut+24b] Jérôme Rutinowski, Simon Klüttermann, Jan Endendyk, Christopher Reining und Emmanuel Müller. „Benchmarking Trust: A Metric for Trustworthy Machine Learning“. In: *2nd World Conference on Explainable Artificial Intelligence (XAI)*. Springer. 2024, S. 287–307. DOI: 10.1007/978-3-031-63787-2\_15.

- [Rut+24c] Jérôme Rutinowski, Hazem Youssef, Sven Franke, Irfan Fachrudin Priyanta, Frederik Polachowski, Moritz Roidl und Christopher Reining. „Semi-automated Computer Vision-based Tracking of Multiple Industrial Entities: A Framework and Dataset Creation Approach“. In: *EURASIP Journal on Image and Video Processing* (2024). DOI: 10.1186/s13640-024-00623-6.
- [RW17] Waseem Rawat und Zenghui Wang. „Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review“. In: *Neural Computation* 29.9 (2017), S. 2352–2449. DOI: 10.1162/NECO\_a\_00990.
- [Rya20] Mark Ryan. „In AI We Trust: Ethics, Artificial Intelligence, and Reliability“. In: *Science and Engineering Ethics* 26.5 (2020), S. 2749–2767. DOI: 10.1007/s11948-020-00228-y.
- [Sag+14] Mohammad A. Saghafi, Aini Hussain, Halimah B. Zaman und Mohamad H. Saad. „Review of Person Re-identification Techniques“. In: *IET Computer Vision* 8.6 (2014), S. 455–474. DOI: 10.1049/iet-cvi.2013.0180.
- [Sak+96] Tomohiko Sakao, Shinsuke Kondoh, Yasushi Umeda und Tet-suo Tomiyama. „The Development of a Cellular Automatic Warehouse“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Bd. 1. 1996, S. 324–331. DOI: 10.1109/IROS.1996.570695.
- [San+18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov und Liang-Chieh Chen. „MobileNetV2: Inverted Residuals and Linear Bottlenecks“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2018, S. 4510–4520. DOI: 10.1109/CVPR.2018.00474.
- [SC21] Anna Severin und Joanna Chataway. „Purposes of Peer Review: A Qualitative Study of Stakeholder Expectations and Perceptions“. In: *Learned Publishing* 34.2 (2021), S. 144–155. DOI: 10.1002/leap.1336.
- [Sch+20] Günther Schuh, Reiner Anderl, Jürgen Gausemeier und Michael ten Hompel, Hrsg. *Industrie 4.0 Maturity Index: Die Digitale Transformation von Unternehmen Gestalten*. 2020. ISBN: 978-3-96834-002-9.
- [Sch+22] Anna Schmitz, Maram Akila, Dirk Hecker, Maximilian Poretschkin und Stefan Wrobel. „An Approach for Systematic Quality Assurance when Working with ML Components“. In: *AT - Automatisierungstechnik* 70.9 (2022), S. 793–804. DOI: doi:10.1515/auto-2022-0012.
- [Sch+23] Nils Schwenzfeier, Jérôme Rutinowski, Marc Hesenius, Christopher Reining und Maribel Acosta. „Generating Embedding Spaces for Re-identifying Pallets from their Chipwood Patterns“. In: *Engineering Applications of Artificial Intelligence*. Elsevier. 2023. DOI: 10.1016/j.engappai.2023.106905.

- [Sch93] Cullen Schaffer. „Selecting a Classification Method by Cross-validation“. In: *Machine Learning* 13 (1993), S. 135–143. DOI: 10.1007/BF00993106.
- [SCK16] Haryong Song, Wonsub Choi und Haedong Kim. „Robust Vision-based Relative-localization Approach using an RGB-depth Camera and LiDAR Sensor Fusion“. In: *IEEE Transactions on Industrial Electronics* 63.6 (2016), S. 3725–3736. DOI: 10.1109/TIE.2016.2521346.
- [SEK21] Jack W. Stokes, Paul England und Kevin Kane. „Preventing Machine Learning Attacks using Authentication and Provenance“. In: *IEEE Military Communications Conference*. 2021, S. 181–188. DOI: 10.1109/MILCOM52596.2021.9653139.
- [Sel+17] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh und Dhruv Batra. „Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization“. In: *IEEE International Conference on Computer Vision*. 2017, S. 618–626. DOI: 10.1109/ICCV.2017.74.
- [Ser+20] Alex Serban, Koen van der Blom, Holger Hoos und Joost Visser. „Adoption and Effects of Software Engineering Best Practices in Machine Learning“. In: *14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 2020, S. 1–12. DOI: 10.1145/3382494.3410681.
- [SGK17] Avanti Shrikumar, Peyton Greenside und Anshul Kundaje. „Learning Important Features through Propagating Activation Differences“. In: *International Conference on Machine Learning*. 2017, S. 3145–3153. DOI: 10.5555/3305890.3306006.
- [SH11] Ahm Shamsuzzoha und Petri T. Helo. „Real-time Tracking and Tracing System: Potentials for the Logistics Network“. In: *International Conference on Industrial Engineering and Operations Management*. 2011, S. 22–24.
- [Sha+13] Ahm Shamsuzzoha, Mikael Ehrs, Richard Addo-Tenkorang, Duy Nguyen und Petri T. Helo. „Performance Evaluation of Tracking and Tracing for Logistics Operations“. In: *International Journal of Shipping and Transport Logistics* 5.1 (2013), S. 31–54. DOI: 10.1504/IJSTL.2013.050587.
- [Sha+19] Shubham Sharma, Shubhankar Verma, Mohit Kumar und Lavanya Sharma. „Use of Motion Capture in 3D Animation: Motion Capture Systems, Challenges, and Recent Trends“. In: *IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*. 2019, S. 289–294. DOI: 10.1109/COMITCon.2019.8862448.
- [SHD21] Morgan K. Scheuerman, Alex Hanna und Emily Denton. „Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development“. In: *ACM Proceedings on Human-Computer Interaction* 5 (2021), S. 1–37. DOI: 10.1145/3476058.

- [She+18] Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen und Xiaogang Wang. „Person Re-identification with Deep Similarity-guided Graph Neural Network“. In: *Proceedings of the European Conference on Computer Vision*. 2018, S. 486–504. DOI: 10.1007/978-3-030-01267-0\_30.
- [Shi+16] Jinqiao Shi, Bingwei He, Liwei Zhang und Jianwei Zhang. „Vision-based Real-time 3D Mapping for UAV with Laser Sensor“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2016, S. 4524–4529. DOI: 10.1109/IRROS.2016.7759666.
- [SHW07] Kristen Stubbs, Pamela J. Hinds und David Wettergreen. „Autonomy and Common Ground in Human-robot Interaction: A Field Study“. In: *IEEE Intelligent Systems* 22.2 (2007), S. 42–50. DOI: 10.1109/MIS.2007.21.
- [SKC21] Charu Sharma, Siddhant R. Kapil und David Chapman. „Person Re-identification with a Locally Aware Transformer“. In: *arXiv preprint arXiv:2106.03720* (2021).
- [SL18] Luc Steels und Ramón López de Mantaras. „The Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe“. In: *AI Communications* 31.6 (2018), S. 485–494. DOI: 10.3233/AIC-180607.
- [Slo88] Jean-Jacques Slotine. „Putting Physics in Control - The Example of Robotics“. In: *IEEE Control Systems Magazine* 8.6 (1988), S. 12–18. DOI: 10.1109/37.9164.
- [SM12] Audie Sumaray und Kami Makki. „A Comparison of Data Serialization Formats for Optimal Efficiency on a Mobile Platform“. In: *6th International Conference on Ubiquitous Information Management and Communication*. 2012, S. 1–6. DOI: 10.1145/2184751.2184810.
- [SM21] Liwei Song und Prateek Mittal. „Systematic Evaluation of Privacy Risks of Machine Learning Models“. In: *30th USENIX Security Symposium*. 2021, S. 2615–2632. ISBN: 978-1-939133-24-3.
- [Soo08] Tan J. Soon. „QR Code“. In: *Synthesis Journal* (2008), S. 59–78.
- [SP07] Uldis Spulle und V. Pusinskis. „Research of Strength of Europallets“. In: *Wood Structure and Properties* 6 (2007).
- [SPP10] Mirko Sokovic, Dusko Pavletic und Kern Pipan. „Quality Improvement Methodologies - PDCA Cycle, RADAR Matrix, DMAIC and DFSS“. In: *Journal of Achievements in Materials and Manufacturing Engineering* 43.1 (2010), S. 476–483. ISSN: 1734-8412.
- [SS02] Daniel Scharstein und Richard Szeliski. „A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms“. In: *International Journal of Computer Vision* 47 (2002), S. 7–42. DOI: 10.1109/SMBV.2001.988771.
- [SS22] Martin Strobel und Reza Shokri. „Data Privacy and Trustworthy Machine Learning“. In: *IEEE Security & Privacy* 20.5 (2022), S. 44–49. DOI: 10.1109/MSEC.2022.3178187.

- [Sta03] Diomidis H. Stamatis. *Failure Mode and Effect Analysis*. Quality Press, 2003. ISBN: 0873895983.
- [Sta19] Diomidis H. Stamatis. *Risk Management Using Failure Mode and Effect Analysis (FMEA)*. Quality Press, 2019. ISBN: 0873899784.
- [Sto+16] Victoria Stodden, Marcia McNutt, David H. Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A. Heroux, John Ioannidis und Michela Taufer. „Enhancing Reproducibility for Computational Methods“. In: *Science* 354.6317 (2016), S. 1240–1241. DOI: 10.1126/science.aah6168.
- [Sto11] John D. Storey. „False Discovery Rate“. In: *International Encyclopedia of Statistical Science* 1 (2011), S. 504–508. DOI: 10.1007/978-3-642-04898-2\_248.
- [Sun+18] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian und Shengjin Wang. „Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline)“. In: *European Conference on Computer Vision*. 2018, S. 480–496. DOI: 10.1007/978-3-030-01225-0\_30.
- [SWM17] Wojciech Samek, Thomas Wiegand und Klaus-Robert Müller. „Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models“. In: *arXiv preprint arXiv:1708.08296* (2017).
- [Sze+15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke und Andrew Rabinovich. „Going Deeper with Convolutions“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2015, S. 1–9. DOI: 10.1109/CVPR.2015.7298594.
- [Szü18] Florian Szücs. „Research Subsidies, Industry-University Cooperation and Innovation“. In: *Research Policy* 47.7 (2018), S. 1256–1266. DOI: 10.1016/j.respol.2018.04.009.
- [Tas00] Gregory Tassej. „Standardization in Technology-based Markets“. In: *Research Policy* 29.4-5 (2000), S. 587–602. DOI: 10.1016/S0048-7333(99)00091-8.
- [TAS18] Mattias Teye, Hossein Azizpour und Kevin Smith. „Bayesian Uncertainty Estimation for Batch Normalized Deep Networks“. In: *International Conference on Machine Learning*. 2018, S. 4907–4916.
- [THH09] Tua A. Tamba, Bonghee Hong und Keum-Shik Hong. „A Path Following Control of an Unmanned Autonomous Forklift“. In: *International Journal of Control, Automation and Systems* 7 (2009), S. 113–122. DOI: 10.1007/s12555-009-0114-y.
- [Tho+17] Martin Thomeier, Ulrike Tagscherer, Elisabeth Schärtl und Andrea Herold. „Future Perspectives: Planning for the Warehouse of the Future“. In: *Swisslog* (2017).
- [Thu22] Bhavani Thuraisingham. „Trustworthy Machine Learning“. In: *IEEE Intelligent Systems* 37.1 (2022), S. 21–24. ISSN: 1541-1672, 1941-1294. DOI: 10.1109/MIS.2022.3152946.

- [Tiw16] Sumit Tiwari. „An Introduction to QR Code Technology“. In: *IEEE International Conference on Information Technology*. 2016, S. 39–44. DOI: 10.1109/ICIT.2016.021.
- [Tjo+21] Erico Tjoa, Hong J. Khok, Tushar Chouhan und Guan Cuntai. „Improving Deep Neural Network Classification Confidence using Heatmap-based Explainable AI“. In: *arXiv preprint arXiv:2201.00009* (2021).
- [TKI17] Toru Takahashi, Yuta Kudo und Rui Ishiyama. „Mass-Produced Parts Traceability System Based on Automated Scanning of ‘Fingerprint of Things’“. In: *15th IAPR International Conference on Machine Vision Applications*. 2017, S. 202–206. DOI: 10.23919/MVA.2017.7986836.
- [TKW10] Ricardo A. Tauro, Benedikt Kaiser und Heinz Wörn. „Path Planning Process Optimization for a Bin Picking System“. In: *41st ISR International Symposium on Robotics and 6th German Conference on Robotics*. 2010, S. 1–7. ISBN: 978-3-8007-3273-9.
- [TLS21] Scott Thiebes, Sebastian Lins und Ali Sunyaev. „Trustworthy Artificial Intelligence“. In: *Electronic Markets* 31 (2021), S. 447–464. DOI: 10.1007/s12525-020-00441-4.
- [Tor+20] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya und Aad van Moorsel. „The Relationship between Trust in AI and Trustworthy Machine Learning Technologies“. In: *ACM Conference on Fairness, Accountability, and Transparency*. 2020, S. 272–283. DOI: 10.1145/3351095.3372834.
- [VB12] Gitte Vanwinckelen und Hendrik Blockeel. „On Estimating Model Accuracy with Repeated Cross Validation“. In: *21st Belgian-Dutch Conference on Machine Learning*. 2012, S. 39–44.
- [Vel02] Manuela M. Veloso. „Entertainment Robotics“. In: *Communications of the ACM* 45.3 (2002), S. 59–63. DOI: 10.1145/504729.504755.
- [Vil+19] Nancy Velásquez Villagrán, Elsa Estevez, Patricia Pesado und Juan De Juanes Marquez. „Standardization: A Key Factor of Industry 4.0“. In: *6th IEEE International Conference on eDemocracy & eGovernment*. 2019, S. 350–354. DOI: 10.1109/ICEDEG.2019.8734339.
- [Viv+10] Kelen Vivaldini, Jorge Galdames, Thales S. Bueno, Roberto C. Araújo, Rafael M. Sobral, Marcelo Becker und Glauco Caubin. „Robotic Forklifts for Intelligent Warehouses: Routing, Path Planning, and Auto-localization“. In: *IEEE International Conference on Industrial Technology*. 2010, S. 1463–1468. DOI: 10.1109/ICIT.2010.5472487.
- [VKB14] Han Van der Maas, Kees-Jan Kan und Denny Borsboom. „Intelligence is what the Intelligence Test Measures. Seriously“. In: *Journal of Intelligence* 2.1 (2014), S. 12–15. DOI: 10.3390/jintelligence2010012.

- [VN16] Robert Varga und Sergiu Nedevschi. „Robust Pallet Detection for Automated Logistics Operations“. In: *International Conference on Computer Vision Theory and Applications*. 2016, S. 470–477. DOI: 10.5220/0005674704700477.
- [Vou+18] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis und Eftychios Protopapadakis. „Deep Learning for Computer Vision: A brief Review“. In: *Computational Intelligence and Neuroscience* (2018). DOI: 10.1155/2018/7068349.
- [Wan+19] Guangcong Wang, Jianhuang Lai, Peigen Huang und Xiaohua Xie. „Spatial-temporal Person Re-identification“. In: *AAAI Conference on Artificial Intelligence*. Bd. 33. 01. 2019, S. 8933–8940. DOI: 10.1609/aaai.v33i01.33018933.
- [WB18] Nicolai Wojke und Alex Bewley. „Deep Cosine Metric Learning for Person Re-identification“. In: *IEEE Winter Conference on Applications of Computer Vision*. 2018, S. 748–756. DOI: 10.1109/WACV.2018.00087.
- [WB52] Norman J. Woodland und Silver Bernard. *Classifying Apparatus and Method*. US Patent 2,612,994. 1952.
- [Wei05] Ron Weinstein. „RFID: A Technical Overview and its Application to the Enterprise“. In: *IT Professional* 7.3 (2005), S. 27–33. DOI: 10.1109/MITP.2005.69.
- [Wei07] Stephen A. Weis. „RFID (Radio Frequency Identification): Principles and Applications“. In: *System* 2.3 (2007), S. 1–23.
- [Wig+18] Benedikt Wigger, Thomas Meissner, Marcel Winkler, Alexander Foerste, Volker Jetter, Andreas Buchholz und André Zimmermann. „Label-/tag-free Traceability of Electronic PCB in SMD Assembly based on Individual Inherent Surface Patterns“. In: *The International Journal of Advanced Manufacturing Technology* 98.9 (2018), S. 3081–3090. DOI: 10.1007/s00170-018-2492-x.
- [Wig20] Benedikt Wigger. „Inhärente Oberflächenstrukturen als Identifikationsmerkmal zur markierungsfreien Einzelteilverfolgung“. In: *Universität Stuttgart*. 2020.
- [Wil+16] Mark D. Wilkinson, Michel Dumontier, IJsbrand J. Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos und Philip E. Bourne. „The FAIR Guiding Principles for Scientific Data Management and Stewardship“. In: *Scientific Data* 3.1 (2016), S. 1–9. DOI: 10.1038/sdata.2016.18.
- [Win21] Jeannette M. Wing. „Trustworthy AI“. In: *Communications of the ACM* 64.10 (2021), S. 64–71. DOI: 10.1145/3448248.
- [Wit18] Frank Witte. *Definition, Historie und Nutzen von Metriken*. Springer, 2018. ISBN: 9783658198442.
- [WMN19] Lukas Wawrla, Omid Maghazei und Torbjørn Netland. „Applications of Drones in Warehouse Operations“. In: *Whitepaper (ETH Zurich, D-MTEC)* (2019).

- [Wu+19] Di Wu, Si-Jia Zheng, Xiao-Ping Zhang, Chang-An Yuan, Fei Cheng, Yang Zhao, Yong-Jun Lin, Zhong-Qiu Zhao, Yong-Li Jiang und De-Shuang Huang. „Deep Learning-Based Methods for Person Re-Identification: A Comprehensive Review“. In: *Neurocomputing* 337 (2019), S. 354–371. DOI: 10.1016/j.neucom.2019.01.079.
- [Wu+20] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang und Yu Philip. „A Comprehensive Survey on Graph Neural Networks“. In: *IEEE Transactions on Neural Networks and Learning Systems* 32.1 (2020), S. 4–24. DOI: 10.1109/TNNLS.2020.2978386.
- [XKM18] Lichao Xu, Vineet R. Kamat und Carol C. Menassa. „Automatic Extraction of 1D Barcodes from Video Scans for Drone-assisted Inventory Management in Warehousing Applications“. In: *International Journal of Logistics Research and Applications* 21.3 (2018), S. 243–258. DOI: 10.1080/13675567.2017.1393505.
- [XRV17] Han Xiao, Kashif Rasul und Roland Vollgraf. „Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms“. In: *arXiv preprint arXiv:1708.07747* (2017).
- [Xu+21] Jiarong Xu, Junru Chen, Siqi You, Zhiqing Xiao, Yang Yang und Jiangang Lu. „Robustness of Deep Learning Models on Graphs: A Survey“. In: *AI Open* 2 (2021), S. 69–78. DOI: 10.1016/j.aiopen.2021.05.002.
- [Ye+21] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao und Steven C. H. Hoi. „Deep Learning for Person Re-identification: A Survey and Outlook“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.6 (2021), S. 2872–2893. DOI: 10.1109/TPAMI.2021.3054775.
- [YHP20] Karen Yeung, Andrew Howes und Ganna Pogrebna. „AI Governance by Human Rights - Centered Design, Deliberation, and Oversight“. In: *The Oxford Handbook of Ethics of AI* (2020), S. 77–106. DOI: 10.1093/oxfordhb/9780190067397.013.5.
- [You+24] Hazem Youssef, Frederik Polachowski, Jérôme Rutinowski, Moritz Roidl und Christopher Reining. „Object Pose Estimation Annotation Pipeline for Multi-view Monocular Camera Systems in Industrial Settings“. In: *Logistics Journal: Referierte Veröffentlichungen* (2024). DOI: 10.2195/lj\_rev\_rutinowski\_en\_202402\_01.
- [Yur+21] Drachuk Yurii, Trushkina Nataliia, Serhieieva Olena, Snitko Yelyzaveta und Belyaeva Ganna. „Standardization of Business Processes in the Transportation Sector“. In: *Open Journal of Business and Management* 9.5 (2021), S. 2312–2328. DOI: 10.4236/ojbm.2021.95125.
- [Zem+13] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi und Cynthia Dwork. „Learning Fair Representations“. In: *30th International Conference on Machine Learning*. Bd. 28. 2013, S. 325–333. DOI: 10.5555/3042817.3042973.

- [ZH08] Huiyu Zhou und Huosheng Hu. „Human Motion Tracking for Rehabilitation — A Survey“. In: *Biomedical Signal Processing and Control* 3.1 (2008), S. 1–18. DOI: 10.1016/j.bspc.2007.09.001.
- [Zha+18] Hui Zhang, Guoliang Shi, Li Liu, Miao Zhao und Zhicong Liang. „Detection and Identification Method of Medical Label Barcode based on Deep Learning“. In: *8th IEEE International Conference on Image Processing Theory, Tools and Applications*. 2018, S. 1–6. DOI: 10.1109/IPTA.2018.8608144.
- [Zha+19] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu und Xindong Wu. „Object Detection with Deep Learning: A Review“. In: *IEEE Transactions on Neural Networks and Learning Systems* 30.11 (2019), S. 3212–3232. DOI: 10.1109/TNNLS.2018.2876865.
- [Zha+20] Xiaofeng Zhao, Rongxu Zhang, Na Zhang, Yanyan Wang, Mingzhou Jin und Shandong Mou. „Analysis of the Shuttle-based Storage and Retrieval System“. In: *IEEE Access* 8 (2020), S. 146154–146165. DOI: 10.1109/ACCESS.2020.3014102.
- [Zhe+15] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang und Qi Tian. „Scalable Person Re-identification: A Benchmark“. In: *IEEE International Conference on Computer Vision*. 2015, S. 1116–1124. DOI: 10.1109/ICCV.2015.133.
- [Zho+18] Yanzhao Zhou, Yi Zhu, Qixiang Ye, Qiang Qiu und Jianbin Jiao. „Weakly Supervised Instance Segmentation using Class Peak Response“. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2018, S. 3791–3800. DOI: 10.1109/CVPR.2018.00399.
- [Zho+19a] Jianlong Zhou, Zhidong Li, Huaiwen Hu, Kun Yu, Fang Chen, Zelin Li und Yang Wang. „Effects of Influence on User Trust in Predictive Decision Making“. In: *CHI Conference on Human Factors in Computing Systems*. 2019, S. 1–6. DOI: 10.1145/3290607.3312962.
- [Zho+19b] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro und Tao Xiang. „Omni-scale Feature Learning for Person Re-identification“. In: *IEEE International Conference on Computer Vision*. 2019, S. 3702–3712. DOI: 10.1109/ICCV.2019.00380.
- [Zho+20] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li und Maosong Sun. „Graph Neural Networks: A Review of Methods and Applications“. In: *Elsevier AI Open* 1 (2020), S. 57–81. DOI: 10.1016/j.aiopen.2021.01.001.
- [ZJ20] Xiao-Ping Zhao und Rui Jiang. „Distributed Machine Learning oriented Data Integrity Verification Scheme in Cloud Computing Environment“. In: *IEEE Access* 8 (2020), S. 26372–26384. DOI: 10.1109/ACCESS.2020.2971519.
- [Zon+20] Tiago Zonta, Cristiano A. da Costa, Rodrigo da Rosa Righi, Miromar J. de Lima, Eduardo Silveira da Trindade und Guann P. Li. „Predictive Maintenance in the Industry 4.0: A Systematic Literature Review“. In: *Computers & Industrial Engineering* 150 (2020). DOI: 10.1016/j.cie.2020.106889.

- [ZYH16] Liang Zheng, Yi Yang und Alexander G. Hauptmann. „Person Re-Identification: Past, Present and Future“. In: *arXiv preprint: 1610.02984* (2016). DOI: 10.48550/arXiv.1610.02984.
- [ZZK05] Wojciech Zajdel, Zoran Zivkovic und Ben J. Krose. „Keeping Track of Humans: Have I Seen this Person Before?“ In: *IEEE International Conference on Robotics and Automation*. 2005, S. 2081–2086. DOI: 10.1109/ROBOT.2005.1570420.

# Anhang

## A.1 Repositories, Datensätze und Dokumentation

- Datensatz A "pallet-block-502:  
<https://zenodo.org/records/6353714>
- Datensatz B "pallet-block-32965:  
<https://zenodo.org/records/6358607>
- Datensatz C "galvanized-636:  
<https://zenodo.org/records/7386956>
- YOLO Datensätze, Modelle und Flugdokumentation für Szenario II:  
<https://zenodo.org/records/11164603>
- Code für die Nutzung des PCB Modells aus Szenario I (Fraunhofer IML Gitlab):  
<https://gitlab.fraunhofer.de/silicon-economy/services/naturalidentificationservice>
- Code für die Nutzung der Ensemblemodelle und des Graphmodells aus Szenario I (Fork von Simon Klüttermann):  
<https://github.com/jerome-rutinowski/PalletEns>
- Code für die Drohnensteuerung aus Szenario II (FLW Github Repository):  
[https://github.com/FLW-TUDO/ReID\\_Drone\\_Scripts](https://github.com/FLW-TUDO/ReID_Drone_Scripts)
- Für die Experimente entwickelte und genutzte Python Library:  
<https://github.com/psorus/eereid>
- Ergebnisse der Expertenbewertungen mittels FRIES Trust Score:  
<https://github.com/jerome-rutinowski/fries-trustscore>

## A.2 Bilder und Icons

Alle verwendeten Icons, die nicht selbst erstellt worden sind, stammen aus Microsoft PowerPoint und den Websites <https://thenounproject.com> sowie <https://flaticon.com>, deren Icons ohne Copyright oder unter Berücksichtigung der Creative Commons Richtlinien verwendet werden können. Alle nicht selbst erstellten Abbildungen sind im jeweiligen Abbildungstext gekennzeichnet.

### A.3 Vorveröffentlichungen und Forschungsprojekte

Zwischenergebnisse dieser Arbeit wurden bereits im Vorfeld publiziert. Die Vorveröffentlichungen sind referenziert und in das Literaturverzeichnis der Arbeit integriert:

- Jérôme Rutinowski, Christian Pionzewski, Tim Chilla, Christopher Reining und Michael ten Hompel. „Towards Re-Identification for Warehousing Entities - A Work-in-Progress Study“. In: *26th IEEE International Conference on Emerging Technologies and Factory Automation*. 2021, S. 501–504. DOI: 10.1109/ETFA45728.2021.9613250
- Jérôme Rutinowski, Hazem Youssef, Anas Gouda, Christopher Reining und Moritz Roidl. „The Potential of Deep Learning based Computer Vision in Warehousing Logistics“. In: *Logistics Journal: Proceedings* (2022). DOI: 10.2195/lj\_proc\_rutinowski\_en\_202211\_01
- Larissa Krämer, Jérôme Rutinowski, Jan Endendyk, Robert Vaut und Moritz Roidl. „On Blockchain-based Token Usage in Cyber-Physical Production Systems“. In: *Logistics Journal: Proceedings* (2022). DOI: 10.2195/lj\_proc\_kraemer\_en\_202211\_01
- Jérôme Rutinowski, Christian Pionzewski, Tim Chilla, Christopher Reining und Michael ten Hompel. „Deep Learning Based Re-identification of Wooden Euro-pallets“. In: *21st IEEE International Conference on Machine Learning and Applications*. 2022, S. 113–117. DOI: 10.1109/ICMLA55696.2022.00023
- Simon Klüttermann, Jérôme Rutinowski, Christopher Reining, Moritz Roidl und Emmanuel Müller. „Towards Graph Representation Based Re-identification of Chipwood Pallet Blocks“. In: *21st IEEE International Conference on Machine Learning and Applications*. 2022, S. 1543–1550. DOI: 10.1109/ICMLA55696.2022.00279
- Jérôme Rutinowski, Bhargav Vankayalapati, Nils Schwenzfeier, Maribel Acosta und Christopher Reining. „On the Applicability of Synthetic Data for Re-Identification“. In: *AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*. 2022. DOI: arXiv:2212.10105
- Nils Schwenzfeier, Jérôme Rutinowski, Marc Hesenius, Christopher Reining und Maribel Acosta. „Generating Embedding Spaces for Re-identifying Pallets from their Chipwood Patterns“. In: *Engineering Applications of Artificial Intelligence*. Elsevier. 2023. DOI: 10.1016/j.engappai.2023.106905
- Jérôme Rutinowski, David Schuster, Sönke Kauffmann, Pedro Arias-Pérez, Frederik Polachowski, Miguel Granero, Moritz Roidl und Pascual Campoy. „Exploring the Re-Identification of Industrial Entities on Autonomous Guided Vehicles“. In: *Logistics Journal: Referierte Veröffentlichungen* (2023). DOI: 10.2195/lj\_proc\_rutinowski\_en\_202310\_01
- Dimitrij-Marian Holm, Philipp Junge, Jérôme Rutinowski und Johannes Fottner. „Investigation of Deep Learning Datasets for Intralogistics“. In: *Conference on Production Systems and Logistics*. 2023. DOI: 10.15488/15311
- Stephan Matzke, Christian Pionzewski, Jérôme Rutinowski, Firdes Celik und Markus König. „Re-identification Robustness Over Time – The Case for Synthetic Training Data“. In: *International Conference on Machine Learning and Applications*. 2023, S. 1363–1370. DOI: 10.1109/ICMLA58977.2023.00206
- Simon Klüttermann, Jérôme Rutinowski, Anh Nguyen, Christopher Reining, Moritz Roidl und Emmanuel Müller. „On Graph Representation based Re-Identification – A Proof of Concept“. In: *IEEE International Conference on Data Mining Workshops (ICDMW)*. 2023, S. 1097–1104. DOI: 10.1109/ICDMW60847.2023.00144

- Jérôme Rutinowski, Hazem Youssef, Sven Franke, Irfan Fachrudin Priyanta, Frederik Polachowski, Moritz Roidl und Christopher Reining. „Semi-automated Computer Vision-based Tracking of Multiple Industrial Entities: A Framework and Dataset Creation Approach“. In: *EURASIP Journal on Image and Video Processing* (2024). DOI: 10.1186/s13640-024-00623-6
- Simon Klüttermann, Jérôme Rutinowski, Anh Nguyen, Britta Grimme, Moritz Roidl und Emmanuel Müller. „On the Effectiveness of Heterogeneous Ensemble Methods for Re-identification“. In: *arXiv preprint arXiv:2403.12606* (2024)
- Jérôme Rutinowski, Simon Klüttermann, Jan Endendyk, Christopher Reining und Emmanuel Müller. „Benchmarking Trust: A Metric for Trustworthy Machine Learning“. In: *2nd World Conference on Explainable Artificial Intelligence (XAI)*. Springer. 2024, S. 287–307. DOI: 10.1007/978-3-031-63787-2\_15

Alle relevanten Publikationen wurden gemäß der Promotionsordnung der Fakultät Maschinenbau der Technischen Universität Dortmund als Vorveröffentlichung indiziert und sind durch das Dekanat anerkannt worden. Die Vorveröffentlichungen sind primär im Rahmen des Forschungsprojekts „Silicon Economy Logistics Ecosystem“ in Zusammenarbeit mit dem Fraunhofer-Institut für Materialfluss und Logistik und als Forschungsarbeiten als Mitglied des „Lamarr-Instituts für Maschinelles Lernen und Künstliche Intelligenz“ entstanden.

#### **A.4 An der Forschungsarbeit beteiligte Personen**

Neben dem Autor waren weitere Personen an Teilen der in dieser Arbeit beschriebenen Forschung beteiligt. Dabei handelt es sich etwa um die gemeinsame Bearbeitung an Forschungsprojekten oder Kollaborationen mit Forschern anderer Universitäten und Forschungsinstitute, namentlich der Technischen Universität München, der Ruhr-Universität Bochum, der Universität Duisburg-Essen, der Universidad Politécnica de Madrid und dem Fraunhofer-Institut für Materialfluss und Logistik. An der Technischen Universität Dortmund wurde spezifisch mit Forschern des Lehrstuhls für Data Science und Data Engineering (Fakultät Informatik) von Prof. Emmanuel Müller und dem Lehrstuhl für Mathematische Statistik und industrielle Anwendungen (Fakultät Statistik) von Prof. Markus Pauly, dem „Research Center Trustworthy Data Science and Security“ und dem „Lamarr-Institut für Maschinelles Lernen und Künstliche Intelligenz“ zusammengearbeitet. Zum Zwecke einer sinnvollen Versuchsplanung, -durchführung und -evaluation wurde das Beratungsangebot des Statistisches Beratungszentrums der Technischen Universität Dortmund wahrgenommen. Diese Personen und Institutionen sowie die Form ihrer Beiträge sind in der untenstehenden Tabelle aufgeführt.

**Tabelle A.1** An der Forschungsarbeit neben dem Autor beteiligte Personen.

<b>Unterstützender Beitrag</b>	<b>Neben dem Autor beteiligte Personen</b>
Implementierung des Graph Modells	Anh Nguyen, Simon Klüttermann
Implementierung des Ensemblemodells	Simon Klüttermann
Implementierung des PCB Modells	Christian Pionzewski, Tim Chilla
Erstellung von Datensatz A	Andrej Plenne, Oleg Belov, Christian Pionzewski, Julian Hinxlage, Tom Stein, Tim Chilla
Erstellung von Datensatz B	Andrej Plenne, Oleg Belov, Christian Pionzewski, Julian Hinxlage, Tom Stein, Tim Chilla
Erstellung von Datensatz C	Jan Endendyk, Bhargav Vankayalapati
Erstellung des YOLO Datensatzes	Frederik Polachowski, Hendrik Aigner
Durchführung von Szenario II	Jan Endendyk, Frederik Polachowski, Hendrik Aigner, Sven Franke, Marie-Claire Olivier
Implementierung der Drohnensteuerung	Frederik Polachowski, Miguel Granero Ramos, David Schuster
Formalisierung des Identifikationsprozesses	Nils Schwenzfeier, Prof. Dr. Maribel Acosta Deibe, Dr. Marc Hesenius
Konzeptionierung des Trust Scores	Simon Klüttermann, Jan Endendyk, Dr.-Ing. Christopher Reining
Beratung hinsichtlich Versuchsdurchführung und statistischer Evaluation	Statistisches Beratungszentrum der Technischen Universität Dortmund
Provision von Euro-Palettenklötzen	European Pallet Association e.V.
Zugang zu NVIDIA DGX-2 Server	Fraunhofer-Institut für Materialfluss und Logistik (IML)
Beratung hinsichtlich Methodik, Vorgehen und Strukturierung des Vorhabens	Dr.-Ing. Christopher Reining

## A.5 Studentische Arbeiten im Kontext der Forschungsarbeit

Während der Erarbeitung dieser Dissertation wurden studentische Abschlussarbeiten betreut, die zum Teil als Vorarbeiten und Anregungen haben dienen können. Alle studentischen Arbeiten, die im Bezug zu dieser Arbeit betreut wurden, sind untenstehend aufgeführt.

**Tabelle A.2** Betreute Abschlussarbeiten im Zusammenhang mit der Dissertation.

Datum	Titel	Typ	Name, Vorname
2023	Lasergestützte Volumenschätzung logistischer Entitäten	BA	Schrötler, Niklas
2022	Finanzielle Potentiale des Trackings von logistischen Entitäten anhand natürlicher Merkmale	PA	Ziehm, Eva Paulina
2022	Mini-drone autonomous missions for indoor warehouse logistics	MA*	Granero Ramos, Miguel
2022	Design of a chipwood re-identification procedure	MA	Kirn, Nikolaus
2022	Development of a procedure for the re-identification of chipwood pallet blocks	MA	Liu, Yanchen
2021	Introducing an automatic pallet track and tracking system in a plant of Pirelli Germany	BA	Lok, Camillus
2021	Einfluss der Entitätserkennung anhand natürlicher Merkmale auf intralogistische Prozesse	BA	Storm, David

PA: Projektarbeit, BA: Bachelorarbeit, MA: Masterarbeit, MA\*: Masterarbeit extern (Universidad Politécnica de Madrid)