Adaptiver Griff-in-die-Kiste – Die methodische Lücke zwischen Forschung und Industrie

T. Stuke ¹, T. Bartsch ², T. Rauschenbach ³

Zusammenfassung

Der Konferenzbeitrag zeigt den Forschungs- und Technikstand bezüglich des Griff-in-die-Kiste auf. Basierend auf einer Literaturrecherche werden Beispiele für regelbasierte und lernende Verfahren vorgestellt. Anschließend erfolgt eine systematische Gegenüberstellung der Verfahren. Hierfür werden die Anforderungen, die ein Griff-in-die-Kiste-System zu erfüllen hat, dargelegt. Die Kriterien resultieren aus einer Expertenbefragung des produktionstechnischen Umfelds der Weidmüller Gruppe. Neben den Anforderungen werden die Gewichtungen zur Bildung einer Rangfolge ermittelt. Die erarbeiteten Anforderungen dienen anschließend zur Bewertung der regelbasierten und lernenden Verfahren. Die Analyse mündet in einer methodischen Lücke zwischen beiden Paradigmen und stellt die Ausgangsbasis für die weitere Arbeit zur Entwicklung des industriellen Griff-in-die-Kiste dar. Abschließend werden erste Arbeitsergebnisse zur Objekterkennung von Reihenklemmen veröffentlicht. In einer Untersuchung werden die Zuverlässigkeit, die Robustheit sowie die Einrichtungsdauer einer Objekterkennung mithilfe von Deep Learning ermittelt. Das angestrebte Forschungsergebnis stellt einen Entwicklungsschritt von automatisierten Systemen, die in einem definierten Wirkbereich eigenständig arbeiten, zu autonomen Systemen, die selbstständig auf zeitvariante Größen reagieren, dar.

Stichwörter

Griff-in-die-Kiste, Bildverarbeitung, Robotik, Deep Learning, lernende Verfahren, regelbasierte Verfahren

1 Motivation

Produzierende Unternehmen stehen in der heutigen Zeit einem Wandel der an sie gestellten Anforderungen gegenüber. Neben den üblichen Zielgrößen wie Kosten, Qualität und Zeit nehmen vor allem Schnelligkeit und Wandlungsfähigkeit an Bedeutung zu, s. [1]. Es ist ein Gesellschaftstrend der Individualisierung zu beobachten, bei dem der Mensch den Wunsch verfolgt, sich von seinen Mitmenschen abzuheben. Als Folge dessen steigt die Produktvielfalt in der Fertigung, während die Losgröße simultan sinkt. Das Resultat in der variantenreichen Produktion elektrotechnischer Komponenten, insbesondere in der Montage, ist bis heute noch oftmals die manuelle Fertigung. Ein großes Hemmnis für die Automatisierung stellt die Bereitstellung der Einzelteile in unstrukturierter Form dar. Die Flexibilität der Hand-Auge-Koordination eines Menschen ist bis heute unübertroffen. Zwar ist die Automatisierung des sog. „Griff-in-die-Kiste“ bereits seit mehreren Jahrzehnten Gegenstand der Forschung, allerdings existiert bis dato keine Lösung, welche die komplexen Anforderungen im industriellen Einsatz ganzheitlich erfüllt.

¹ Tobias Stuke, M. Sc., Weidmüller Interface GmbH & Co. KG, Detmold

² Prof. Dr.-Ing. Thomas Bartsch, TH Ostwestfalen-Lippe, Lemgo

³ Prof. Dr.-Ing. habil. Thomas Rauschenbach, Fraunhofer IOSB-AST, Ilmenau

2 Stand der Technik

Mit Beginn der 1990er-Jahre untersuchen Ferrari und Canny [2] die Prozessgrößen des automatisierten Greifens von Bauteilen. Unter Verwendung der Normalkraft sowie des Reibungskoeffizienten berechnet sich die Coulombsche Reibung. Diesen physikalischen Effekt, auch Kraftschluss genannt, nutzen Ferrari und Canny [2] zur Beschreibung des sog. „Wrench“. Das Wrench berechnet sich aus der euklidischen Norm der äußeren Kräfte und Momente, die auf das Objekt wirken, ohne dass der Kraftschluss unterbricht, s. [2]. Die analytische Größe wird im Weiteren dazu genutzt, um geeignete Greifpunkte von ungeeigneten zu unterscheiden, s. [3]. Dupuis et al. [3] nutzen einen Algorithmus mit definierter Schrittweite zur Abtastung diverser Greifpunkte eines einzelnen Pleuels entlang der Richtungsvektoren des dreidimensionalen Raums. Die abgetasteten Greifpunkte werden hinsichtlich des Wrench bewertet und deren Anzahl durch die Anwendung eines Schwellwertes reduziert.

Im weiteren Forschungsverlauf findet die Beschreibung der Objekte mithilfe von CAD-Modellen zunehmend Verwendung, s. [4–8]. Heutzutage haben sich daraus kommerzielle Lösungen, z.B. der Hersteller Keyence, Photoneo oder Pickit, entwickelt, s. [9–11]. Mithilfe des 3D-Modells werden durch den Anwender Erkennungsmerkmale, wie z.B. Ecken und Kanten, aber auch mögliche Greifpunkte im Vorfeld definiert. Die Merkmale dienen anschließend dazu, das Objekt im Ladungsträger zu identifizieren und zu lokalisieren. Die Objektlokalisierung ist eine elementare Aufgabe, für die Buchholz et al. den sog. „RANSAM“-Algorithmus [4], engl. für „Random Sample Matching“, entwickelten. Ziel des Algorithmus ist die Ermittlung einer validen Pose im Haufwerk, unter Anwendung einer schrittweisen Translation und Rotation des Objektes. Spenrath und Pott [5] nutzen für die Ermittlung der Objekt-Pose hingegen ein heuristisches Suchverfahren.

Die zuvor beschriebenen Verfahren werden im Weiteren als „regelbasierte Verfahren“ bezeichnet. Darüber hinaus existiert eine weitere Klasse an Verfahren, die auf Methoden des maschinellen Lernens, wie z.B. dem Deep Learning basieren, nachfolgend „lernende Verfahren“ genannt. Die grundlegende Unterteilung der Verfahren geht auf die Veröffentlichung von Sahbani et al. [12] zurück, in der die Notation „analytische“ und „datengetriebene“ Verfahren Verwendung findet.

Lernende Verfahren nutzen ein Faltungsnetzwerk, engl. „Convolutional Neural Network“, zur Objektlokalisierung oder Greifpunktermittlung. Anhand von Tiefenbildern wird der Zusammenhang zwischen Bildpixel und Greifqualität trainiert, s. [13–16]. Im Beispiel von Pinto und Gupta [13] werden 50.000 Entnahmezyklen zum Training des neuronalen Netzes durchlaufen. Dies entspricht einer Lerndauer von ca. 700 Stunden. Wissenschaftler des Unternehmens Google führen ein Experiment durch, den Griff-in-die-Kiste durch Versuch und Irrtum in unüberwachter Form zu erlernen. In der Studie werden dazu 14 Roboter in 800.000 Lernzyklen über eine Lerndauer von zwei Monaten trainiert, s. [15]. Mahler et al. [14] führen den Lernvorgang ihres „Dex-Net 2.0“-Modells hingegen in überwachter Form durch. Hierfür kommen 6,7 Mio. Tiefenbilder, welche mit guten und schlechten Greifpunkten annotiert sind, zum Einsatz. Die Wissenschaftler Dong et al. [17], Xiang et al. [18] und Deng et al. [19] verfolgen wiederum den Ansatz, die sechs Freiheitsgrade eines räumlich beliebig angeordneten Objektes zu erlernen. Aufgrund der umfangreichen Datenmengen und zahlreichen Trainingszyklen, die für einen Lernprozess vonnöten sind, nehmen in der heutigen Zeit Simulationen stark an Bedeutung zu. Den Einsatz von Simulationstechniken untersuchen Shao et al. [20], Bousmalis et al. [21] und Sundermeyer et al. [22]. Letztere nutzen die Simulation dazu, Greifsznarien anhand der Objektpunktwolke zu erzeugen, welche im Anschluss hinsichtlich der Greifqualität bewertet werden. Der Einsatz von Simulationen ermöglicht somit die Erzeugung sehr großer Datensätze, welche zum Training der neuronalen Netze verwendet werden. Der Dex-Net 2.0-Ansatz von Mahler et al. [14] verdeutlicht dieses Vorgehen. Aktuelle Forschungsansätze thematisieren überdies die Verwendung eines geschlossenen Regelkreises zur Bahnführung des Roboters durch das Bildverarbeitungssystem, s. [23, 24]. Diese haben den Vorteil, dass die Trajektorie des Werkzeugmittelpunktes in Echtzeit durch das Bildverarbeitungssystem geregelt wird. In einem Versuch demonstrieren Morrison et al. [23], dass der Roboter auf eine Translation des Objekts um 25 mm sowie eine Rotation um 10 ° dynamisch reagiert, nachdem die Roboterbewegung bereits eingeleitet hat.

3 Verfahrensvergleich

In diesem Abschnitt werden die regelbasierten und lernenden Verfahren miteinander verglichen und bewertet. Daraus leitet sich eine methodische Lücke ab, welche es zu minimieren gilt, um den Griff-in-die-Kiste im industriellen Einsatz zu etablieren.

3.1 Anforderungen

Im Rahmen der Verfahrensanalyse ist festzustellen, dass keine einheitlichen Bewertungskriterien für Griff-in-die-Kiste-Systeme existieren. Im Rahmen einer Expertenbefragung wurden deshalb Anforderungen an den industriellen Griff-in-die-Kiste innerhalb der Weidmüller Gruppe erarbeitet, s. [25]. Der Teilnehmerkreis setzt sich aus den Fachbereichen Fertigungsplanung, Betriebsmittelplanung, Vorentwicklung, Steuerungsbau, Konstruktion sowie der Montagetechnik zusammen. Die resultierende Anforderungsliste ist Tabelle 1 dargestellt.

Tabelle 1: Anforderungen an den Griff-in-die-Kiste

Nr.	Anforderung	Min. Wert	Max. Wert	Zielwert	Einheit
AF-1	Universalität	3	∞	4	Anzahl der Varianten
AF-2	Ordnungszustand	OZ = 0 / 0	OZ = 3 / 3	OZ = 0 / 0	OZ = OG / PG
AF-3	Objektgröße	5	150	75	Länge in mm
AF-4	Herstellerunabhängigkeit	-	-	hoch	-
AF-5	Einrichtdauer	-	5	4	Zeit in h
AF-6	Zykluszeit	-	12	10	Zeit in s
AF-7	Zuverlässigkeit	80	100	85	Wahrscheinlichkeit in %
AF-8	Robustheit	-	-	hoch	-
AF-9	Kontextwissen	-	-	hoch	-
AF-10	Positioniergenauigkeit	1	8	4	Abstand in mm

Im Nachfolgenden erfolgt eine Erläuterung der Anforderungen. AF - 1: Die Universalität beschreibt die Anpassungsfähigkeit des Systems. Die Implementierung einer neuen Variante ist im Idealfall ohne Aufwand, wie z.B. mechanische Umbauten oder Programmierarbeiten, durchführbar. AF - 2: Der Ordnungszustand (OZ) beschreibt den Orientierungsgrad (OG) und Positionierungsgrad (PG) eines Bauteils, s. [26]. Die Anforderung an das Handhabungssystem sind umso komplexer, je geringer der Ordnungszustand ist. AF - 3: Die Objektgröße nimmt Einfluss auf die Komplexität der technischen Handhabungslösung. In dieser Arbeit wird die Ausprägung einer geringen Objektgröße fokussiert, da die zu verarbeitenden elektrotechnischen Komponenten klein sind. AF - 4: Die Herstellerunabhängigkeit beschreibt die Möglichkeit, mit unternehmensinternen Ressourcen Erweiterungen oder Anpassungen an der Hardware oder Software des Systems vorzunehmen. AF - 5: Die Einrichtdauer umfasst den Zeitaufwand, um das Handhabungssystem auf einen bis dato unbekannten Artikel vorzubereiten. Tätigkeiten, die im Zuge des Einrichtens anfallen, sind z.B. mechanisches Umrüsten des Werkzeugs oder programmtechnische Anpassungen. AF - 6: Die Zykluszeit umfasst die Dauer einer vollständigen Handhabungsabfolge, von der Objekterkennung bis zur Objektbereitstellung. AF - 7: Die Zuverlässigkeit entspricht der statistischen Wahrscheinlichkeit, mit der die Teile dem nachfolgenden Prozessschritt zur Verfügung gestellt werden. Sie berechnet sich anhand des Quotienten aus der Anzahl an korrekt entnommenen Objekten und der Anzahl an Entnahmeversuchen. AF - 8: Die Robustheit bezeichnet die Fähigkeit eines Systems, Veränderungen, ohne Anpassung der anfangs stabilen Konfiguration, zu widerstehen, s. [27]. Ein Beispiel für eine zeitvariable Störgröße ist die Belichtungsintensität, welche in Abhängigkeit des Tageslichteinfalls variiert. AF - 9: Das Kontextwissen stellt den übergeord-

neten Rahmen der Handhabungsaufgabe dar. Dieses umfasst neben der Information über die Entnahme-Pose des Objektes auch die Information bezüglich dessen Bereitstellungs-Pose, um den nachgeschalteten Prozessschritt zu ermöglichen. AF - 10: Die Positioniergenauigkeit beschreibt die absolute Genauigkeit, mit der das Handhabungsobjekt nach erfolgreicher Entnahme aus dem Ladungsträger bereitgestellt wird.

Die genannten Anforderungen besitzen eine unterschiedliche Relevanz hinsichtlich der Griff-in-die-Kiste-Lösung, weshalb im Weiteren die Gewichtung der Anforderungen erläutert wird. Im Rahmen der o.g. Expertenbefragung wurden die einzelnen Anforderungen dazu einem paarweisen Vergleich unterzogen. Die resultierende Rangfolge der Anforderungen ist in Abbildung 1 dargestellt.

Abbildung 1: Rangfolge der Anforderungen an den Griff-in-die-Kiste

Das Resultat verdeutlicht, dass die wichtigsten Anforderungen an den Griff-in-die-Kiste eine hohe Zuverlässigkeit (18,3 %) sowie eine hohe Robustheit (17,1 %) sind. Die Gewährleistung eines stabilen Teilestroms ist von höchster Priorität, um die nachfolgenden Produktionsschritte sicherzustellen. Die Anforderungen mit dritthöchster Priorität sind die Universalität (14,6 %) und das Kontextwissen (14,6 %). Demnach gilt es eine Lösung zu entwickeln, die für diverse Artikelvarianten einsetzbar ist und die neben dem Wissen über die Entnahmeaufgabe auch Kenntnis über den nachfolgenden Prozessschritt und der dafür benötigten Bereitstellungs-Pose besitzt. Im Feld der mittleren Wertigkeit befinden sich die Anforderungen Einrichtdauer (12,2 %) und Zykluszeit (9,8 %). Die Bedeutung der Kriterien Ordnungszustand (6,1 %), Herstellerunabhängigkeit (3,7 %) und Positioniergenauigkeit (3,7 %) ist gering. Die Objektgröße (0,0 %) wird als unwichtig erachtet und scheidet gänzlich aus. Die Anforderungen werden im Weiteren zur Beurteilung der Griff-in-die-Kiste-Verfahren verwendet.

3.2 Verfahrensbewertung

In diesem Abschnitt findet eine kritische Reflektion der vorhandenen Griff-in-die-Kiste-Ansätze (Kap. 2) statt. Für die Beurteilung der regelbasierten und lernenden Verfahren wird eine Nutzwertanalyse herangezogen. Es gilt die Frage zu beantworten, wie gut die erarbeiteten Anforderungen durch die jeweilige Methode erfüllt werden. Die Bewertung erfolgt anhand einer dreistufigen Skala:

- Bewertungszahl 0: Anforderung nicht erfüllt,
- Bewertungszahl 0,5: Anforderung bedingt erfüllt und
- Bewertungszahl 1: Anforderung überwiegend erfüllt.

Der Erfüllungsgrad der Anforderungen berechnet sich aus dem Produkt der Bewertungszahl und des Gewichtungsfaktors. Die Bewertung der regelbasierten Verfahren wird anhand der Literaturstellen [3–8, 28] durchgeführt. Zur Bewertung der lernenden Verfahren werden die Ansätze [14–16, 20, 23, 24, 29–31] herangezogen. Das Bewertungsergebnis ist in Abbildung 2 visualisiert und zeigt die Stärken beider Verfahrenstypen. Gleichzeitig wird die Anforderungsgewichtung einbezogen, sodass die Kriterien mit der höchsten Priorität im Fokus stehen. Die Stärken der lernenden Verfahren sind allen voran die Universalität sowie die Robustheit. In den Experimenten von Mahler et al. [14] und Levine et al. [15] werden Entnahmevorgänge von diversen Alltagsgegenständen aus einem Behälter durchgeführt. Durch den gelernten Zusammenhang zwischen Bildpixel und Greifpunkt ist das System in der Lage beliebige Objekte zu greifen. Regelbasierte Verfahren besitzen hingegen eine sehr eingeschränkte Universalität. Lediglich exakt definierte Teile werden einwandfrei identifiziert, lokalisiert und gegriffen. In Bezug auf das Kontextwissen kehren sich die Stärken der regelbasierten und lernenden Verfahren um. Regelbasierte Verfahren verfügen generell über Kontextwissen, da dieses durch den Anwender, in Form der gewünschten Ablage-Pose, programmiert wird. In dem Anwendungsfall von Ellekilde und Petersen [7] werden Stahlwellen nach der Entnahme aus dem Behälter in eine Presse eingelegt. Die entsprechende Ziel-Pose wird dabei manuell definiert, s. [7].

Abbildung 2: Stärken-Diagramm der Griff-in-die-Kiste-Verfahren

Lernende Verfahren verfügen bis dato nicht über Kontextwissen. Aufgrund der Tatsache, dass lediglich der Zusammenhang zwischen Bildpixel und Objekt-Pose bzw. Greifpunkt gelernt wird, hat das Modell keine Kenntnis darüber, welches Teil gegriffen wird oder welcher Prozessschritt im Anschluss folgt. Häufig wird das Objekt nach erfolgreicher Entnahme außerhalb des Behälters fallen gelassen, s. [14, 15]. Eine Stärke der lernenden Verfahren gegenüber den regelbasierten Verfahren ist die Robustheit. Lernende Verfahren sind in der Lage, verschiedene Belichtungen und Hintergründe zu erlernen, sofern der Trainingsdatensatz eine entsprechende Varianz besitzt. Durch die sensorische Aufnahme des Behälterinhaltes aus 50 verschiedenen Perspektiven, auch Mehrfachansicht genannt, gelingt Morrisson et al. [23] eine Steigerung der Greifqualität um 10 %. Im Zuge dessen ist ein Anstieg der durchschnittlichen Dauer für einen Greifzyklus von 8,8 s auf 11,4 s zu verzeichnen. Darüber hinaus wird ein geschlossener Regelkreis zur Bahnführung des Roboters unter Verwendung des Sensorsignals angewandt. Dadurch kann das System auf ein Verrutschen der Objekte im Behälter reagieren.

Maßnahmen wie diese erhöhen die Robustheit eines Griff-in-die-Kiste-Systems. Regelbasierte Verfahren verfügen hingegen über ein Sensorsystem, welches speziell auf die vorherrschenden Bedingungen abgestimmt ist. Varianz in der Umgebungsbelichtung oder in der Hintergrundhelligkeit wirkt sich negativ auf die Zuverlässigkeit der Teileerkennung aus. Infolgedessen werden zusätzliche Maßnahmen zur Sicherstellung einer robusten Objekterkennung erforderlich, wie das Anwendungsbeispiel des Unternehmens BMW verdeutlicht. Für die Handhabung von Karosserieteilen werden Gitterboxen speziell mit einem pinkfarbenen Kunststoff zur Erzeugung eines kontrastreichen, invarianten Hintergrunds ausgestattet, s. [32]. Eine Schwäche sowohl der regelbasierten als auch der lernenden Verfahren ist die hohe Einrichtdauer. Für das maschinelle Lernen des Griff-in-die-Kiste ist eine hohe Anzahl an Lernzyklen erforderlich, welche in physischen Tests oder Simulationen generiert werden. Im Beispiel von Deng et al. [19] beträgt die Trainingsdauer für 20 verschiedene Objekte 12 Stunden, s. [19, S. 3669]. Darüber hinaus ist eine Annotation der Trainingsdaten im Fall des überwachten Lernens erforderlich, welche einen zusätzlichen Zeitaufwand darstellt, s. [14]. Regelbasierte Verfahren sind ebenfalls zeitintensiv in der Initialisierung der Objekte, da Erkennungsmerkmale sowie Greifpunkte manuell durch den Anwender mithilfe des 3D-Modells zu definieren sind. Der Hersteller Photonéo demonstriert anhand eines Videos, dass das Einrichten eines Objektes für eine erfahrene Person innerhalb 30 Minuten realisierbar ist, s. [10]. Für das Anwendungsbeispiel von Deng et al. [19] entspricht dies einer Einrichtdauer von 10 Stunden. Beide Verfahren überschreiten somit die Anforderung an die Einrichtdauer von max. 5 Stunden, s. Abschn. 3.1. Die Zuverlässigkeit der regelbasierten Verfahren ist geringfügig höher als die der lernenden Verfahren.

3.3 Methodische Lücke

Regelbasierte und lernende Verfahren weisen individuelle Stärken auf, s. Abbildung 2. Der grüne Linienzug kennzeichnet ein ideales Verfahren, welches alle Anforderungen vollständig erfüllt. Kein bis dato vorhandenes Verfahren wird den Anforderungen des industriellen Griff-in-die-Kiste vollumfänglich gerecht, was auf einen Zielkonflikt zwischen den Anforderungskombinationen zurückzuführen ist. In Abbildung 3 ist der Zielkonflikt zwischen Kontextwissen und Universalität dargestellt.

Abbildung 3: Zielkonflikt zwischen Kontextwissen und Universalität

Regelbasierte Systeme verfügen zwar über Kontextwissen, sind jedoch durch eine geringe Universalität gekennzeichnet. Sie werden heutzutage in spezialisierten Anwendungsfällen, wie z.B. in der Automobilindustrie zur Handhabung großvolumiger Teile, eingesetzt.

Das Stärken-Verhältnis der lernenden Verfahren in Bezug auf Kontextwissen und Universalität ist gegenüber den regelbasierten Systemen invertiert. Dadurch finden sie überwiegend in der Forschung Anwendung, welche z.B. die Handhabung von diversen Alltagsgegenständen fokussiert. In Abbildung 3 sind beide Paradigmen hinsichtlich der Erfüllung von Kontextwissen und Universalität eingeordnet. Es zeichnet sich ein Zielkonflikt zwischen den gegensätzlichen Anforderungen ab, welcher zur Spaltung der regelbasierten und lernenden Verfahren führt. Die daraus resultierende methodische Lücke ist im oberen, rechten Quadranten der Abbildung 3 dargestellt. Zum heutigen Stand existiert keine Lösung, welche die Stärken beider Paradigmen kombiniert. Resümierend gilt es für den industriellen Einsatz des Griff-in-die-Kiste eine hybride Automatisierungslösung zu entwickeln, die folgende Anforderungen simultan erfüllt:

- hohe Zuverlässigkeit,
- hohe Robustheit,
- hohes Kontextwissen,
- niedrige Einrichtdauer.

4 Lösungsansatz

Aus dem Verfahrensvergleich (Kap. 3) geht hervor, dass das Deep Learning einen geeigneten Lösungsansatz für die flexible Objekterkennung darstellt. Im Weiteren wird die Untersuchung der Anwendbarkeit eines Faltungsnetzwerks zur Klassifikation und Lokalisierung von elektrotechnischen Komponenten im Rahmen einer Studie der Weidmüller Gruppe erläutert.

Ziel der Untersuchung ist es, eine Einschätzung über die Zuverlässigkeit, die Robustheit sowie die Einrichtdauer eines Faltungsnetzwerks zur Erkennung von elektrotechnischen Komponenten zu erhalten. In der Versuchsdurchführung werden vier Reihenklemmen und ein Steckverbinder auf einer Tischebene angeordnet, ohne dass eine Überlappung der Teile auftritt. Zur Objekterkennung wird die „SSD Mobilenet V2“-Architektur mithilfe des Programmiergerüsts „TensorFlow“ von Google genutzt, s. [33]. Das Faltungsnetzwerk ist über die Plattform „GitHub“ [34] öffentlich zugänglich und stellt eine Weiterentwicklung des sog. „Single Shot Detector (SSD)“-Modells [35] dar, welches nach dem Funktionsprinzip der einstufigen Objekterkennung mittels exemplarweiser Segmentierung arbeitet. Die Architektur wurde bereits anhand des „COCO“-Datensatzes von GitHub-Nutzern trainiert. COCO ist ein Akronym für „Common Objects in Context“, s. [36]. Der Datensatz enthält 328.000 Bilder und bildet 1,5 Mio. Alltagsgegenstände im üblichen Kontext ab. Zur Ausführung des SSD Mobilenet V2 werden zwei verschiedene Systeme herangezogen: a) ein Laptop und b) ein PC. Diese unterscheiden sich hauptsächlich durch die Grafikleistung voneinander. Der Laptop verfügt über keine separate Grafikeinheit, wohingegen der PC über eine „GeForce RTX 2080“ verfügt.

Die Versuchsdurchführung unterteilt sich in zwei Phasen: a) der Trainingsphase und b) der Testphase. Während der Trainingsphase gilt es das Netzwerk für die Klassifikation elektrotechnischer Komponenten anzupassen. Die Modifikation, auch Fine-Tuning genannt, erfolgt anhand eines annotierten Datensatzes, welcher 205 Bilder beinhaltet und im Unternehmen Weidmüller erstellt wurde. Die Bilder zeigen die Komponenten aus diversen Perspektiven, unterschiedlichen Distanzen sowie mit verschiedenen Hintergründen. Für die Erstellung der Testbilder sowie deren Annotation mit sog. „Labels“ und „Bounding Boxes“ werden in Summe zwei Stunden benötigt. Der anschließende Trainingsvorgang wird zum Vergleichszweck sowohl mit dem Laptop als auch mit dem PC durchgeführt. Die Lerndauer unter Verwendung des Laptops beträgt 1 Stunde und 30 Minuten, die Trainingsdauer unter Verwendung des PCs sieben Minuten.

Nach Abschluss des Trainings erfolgt die Testphase, in welcher die Objekterkennung mithilfe eines zeitkontinuierlichen Bildsignals durchgeführt wird. Währenddessen werden die Position und Orientierung der Komponenten variiert, um die Robustheit und Zuverlässigkeit zu prüfen. Des Weiteren erfolgt eine Veränderung der Belichtungsintensität durch Steuerung des Lichteinfalls eines Fensters mithilfe einer Jalousie. Ebenfalls wird die Hintergrundfarbe durch Verschieben der Teile von einer hellen Tischoberfläche auf eine schwarze Tastatur verändert. Der Versuch zeigt, dass die Objekterkennung eine hohe Robustheit aufweist. Das Objekt wird trotz Translation und Rotation größtenteils korrekt klassifiziert und lokalisiert. Durch die Veränderung der Belichtungsintensität reduziert sich die Zuverlässigkeit von 98 % bei schwacher Belichtung lediglich auf 97 % bei starker Belichtung. Beim Wechsel des Hintergrunds von hell auf dunkel sinkt hingegen die Zuverlässigkeit der Objekterkennung von ca. 80 % auf ca. 70 %. Die Helligkeit des Hintergrunds nimmt somit einen größeren Einfluss auf die Robustheit als die Belichtungsintensität. Darüber hinaus zeigt der Test auch die Grenzen der Zuverlässigkeit auf. Wird die Reihenklammer z.B. hochkant auf die Tischoberfläche gestellt, ist eine Erkennung nicht möglich. Des Weiteren ist die Erkennung einer beigen Reihenklammer, unter ansonsten ausschließlich grünen Reihenklammern, nur mit geringer Zuverlässigkeit von etwa 60 % möglich. Die Trainingsphase spiegelt den Einrichtbetrieb der Objekterkennung eines Griff-in-die-Kiste-Systems wider. Unter Verwendung des Laptops beträgt die Einrichtdauer 3 Stunden und 30 Minuten. Die Einrichtdauer unter Verwendung des PCs beträgt hingegen 2 Stunden und 7 Minuten. Die signifikante Zeitreduktion um ca. 40% ist der höheren Grafikleistung des PCs gegenüber dem Laptop zuzuschreiben. Beide Szenarien erfüllen die Anforderung an die Einrichtdauer von 4 Stunden, s. Abschn. 3.1. Die Testphase zeigt, dass die Varianz des Trainingsdatensatzes einen hohen Einfluss auf die Zuverlässigkeit und die Robustheit der Objekterkennung nimmt. Es ist von großer Bedeutung, dass der Datensatz Bildaufnahmen der Objekte aus verschiedenen Positionen, Entfernungen, Belichtungen und Hintergründen enthält. Für die Zukunft gilt es die Zuverlässigkeit der Objekterkennung zu optimieren. Dies kann zum einen durch die Vergrößerung der Eingangsschicht des Faltungsnetzwerks von 320 x 320 Bildpixel auf 640 x 640 Bildpixel erfolgen oder zum anderen durch eine Erweiterung des Trainingsdatensatzes um Detailaufnahmen der Komponenten.

5 Fazit und Ausblick

Die Analyse des Technikstands (Kap. 2) zeigt die historischen sowie die gegenwärtigen Ansätze des Griff-in-die-Kiste auf. Daraus leiten sich zwei grundlegende Automatisierungsparadigmen ab: a) die regelbasierten Verfahren und b) die lernenden Verfahren. Im Rahmen der Expertenbefragung (Kap. 3) wurden 10 Anforderungen an eine zukünftige Lösung des industriellen Griff-in-die-Kiste erarbeitet und gewichtet. Die Bewertung beider Paradigmen verdeutlicht, dass weder die regelbasierten Verfahren noch die lernenden Verfahren eine geeignete Lösung darstellen. Die methodische Lücke zeichnet sich anhand des Zielkonflikts der Anforderungskombinationen, wie z.B. der Universalität und des Kontextwissens, ab. Es besteht der Bedarf an einem hybriden Verfahren, welches die Stärken beider Paradigmen vereint und die methodische Lücke schließt. Der Fokus in der Entwicklung des Griff-in-die-Kiste-Systems ist gemäß Anforderungsgewichtung auf die Zuverlässigkeit, die Robustheit, das Kontextwissen und die Einrichtdauer zu legen. Einen Lösungsansatz stellt hierbei die Bildverarbeitung unter Verwendung des Deep Learnings dar. Die Versuchsdurchführung (Kap. 4) zur Klassifikation und Lokalisierung elektrotechnischer Komponenten zeigt, dass eine robuste und universelle Objekterkennung mithilfe eines Faltungsnetzwerks realisierbar ist. Hierbei ist es von großer Bedeutung, dass die Bilder des Trainingsdatensatzes eine hohe Varianz in Bezug auf Perspektive, Belichtung und Hintergrund aufweisen. In Bezug auf die Einrichtdauer ist die Grafikleistung des verwendeten Systems zum Training des neuronalen Netzes entscheidend.

Für den weiteren Entwicklungsverlauf gilt es die Zuverlässigkeit der Objekterkennung zu optimieren, um z.B. die Erkennung aufrechtstehender Teile zu gewährleisten. Zielführend ist die Vergrößerung

der Netzwerkeingabeschicht sowie die Erweiterung des Trainingsdatensatz um zusätzliche Detailaufnahmen der Produkte. Darüber hinaus gilt es die flexible Objekterkennung von chaotisch bereitgestellten Teilen zu untersuchen. Für diese Aufgabe ist ein erhöhter Schwierigkeitsgrad zu erwarten, da sich wiederkehrende Objektmerkmale verschiedener Komponenten überlappen. Einen potenziellen Lösungsansatz stellt in diesem Zusammenhang die Verwendung der Tiefeninformation des Sensors sowie die Mehrfachansicht des Behälterinhaltes dar. Hinsichtlich der Optimierung der Robustheit des Greifvorgangs ist die Nutzung eines geschlossenen Regelkreises für die Regelung der Trajektorie geplant.

6 Literatur

- [1] J. Bischoff et al., *Fabriken im Wandel: Gestaltungsprinzipien der Veränderungsfähigkeit und ihre Vorteile für Produktionsunternehmen*. [Online]. Verfügbar unter: https://www.agiplan.de/wp-content/uploads/2019/05/Fabriken_im_Wandel_2016.pdf (Zugriff am: 11. Dezember 2021).
- [2] C. Ferrari und J. Canny, „Planning optimal grasps“ in *1992 International Conference on Robotics and Automation Nice, France - May 1992*.
- [3] D. Dupuis, S. Leonard, M. Baumann, E. Croft und J. Little, „Two Fingered Grasp Planning for Randomized Bin Picking“ in *Proc. of Robotics, Science and Systems 2008 Manipulation Workshop*.
- [4] D. Buchholz, M. Futterlieb, S. Winkelbach und F. Wahl, *2013 IEEE International Conference on Robotics and Automation (ICRA 2013): Karlsruhe, Germany, 6 - 10 May 2013*. Piscataway, NJ: IEEE, 2013.
- [5] F. Spenrath und A. Pott, „Gripping Point Determination for Bin Picking Using Heuristic Search“, *Procedia CIRP*, Jg. 62, S. 606–611, 2017, doi: 10.1016/j.procir.2016.06.015.
- [6] C. Martinez, H. Chen und S. Nidamarthi, „Automated bin picking system for randomly located industrial parts“ in *2015 IEEE International Conference on Technologies for Practical robot Applications (TePRA)*, Woburn, MA, USA, 2015, S. 1–6, doi: 10.1109/TePRA.2015.7219656.
- [7] L.-P. Ellekilde und H. Petersen, „Motion planning efficient trajectories for industrial bin-picking“, *The International Journal of Robotics Research*, Jg. 32, 9-10, S. 991–1004, 2013, doi: 10.1177/0278364913487237.
- [8] D. Kraft, L.-P. Ellekilde und J. Jørgensen, „Automatic Grasp Generation and Improvement for Industrial Bin-Picking“ in *Springer Tracts in Advanced Robotics, Gearing Up and Accelerating Cross-fertilization between Academic and Industrial Robotics Research in Europe*, F. Röhrbein, G. Veiga und C. Natale, Hg., Cham: Springer International Publishing, 2014, S. 155–176, doi: 10.1007/978-3-319-03838-4_8.
- [9] Keyence Deutschland GmbH, *3D Robot Vision CV-X480D*. Produktbroschüre. [Online]. Verfügbar unter: https://www.keyence.de/landing/vision/pr_3d_robot_vision.jsp (Zugriff am: 11. Dezember 2021).
- [10] Photoneo s.r.o., *Bin Picking Studio: Die vielseitigste Roboter-Intelligenz-Software*. [Online]. Verfügbar unter: <https://www.photoneo.com/de/bin-picking-studio-alles-in-einem-toolbox-fuer-roboterintegratoren/> (Zugriff am: 11. Dezember 2021).
- [11] Pick-it N.V., *Pickit 3D-Produkte - Pickit 3D: Pickit M-HD, Pickit L-HD, Pickit L*. [Online]. Verfügbar unter: <https://www.pickit3d.com/de/produkte> (Zugriff am: 11. Dezember 2021).
- [12] A. Sahbani, S. El-Khoury und P. Bidaud, „An overview of 3D object grasp synthesis algorithms“ (en), *Robotics and Autonomous Systems*, Jg. 60, Nr. 3, S. 326–336, 2012, doi: 10.1016/j.robot.2011.07.016.
- [13] L. Pinto und A. Gupta, „Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours“ in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, 2016, S. 3406–3413.
- [14] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan und X. Liu, „Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics“, 27. März 2017. [Online]. Verfügbar unter: <http://arxiv.org/pdf/1703.09312v3>.
- [15] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz und D. Quillen, „Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection“, *The International Journal of Robotics Research*, Jg. 37, 4-5, S. 421–436, 2018, doi: 10.1177/0278364917710318.
- [16] A. Mousavian, C. Eppner und D. Fox, „6-DOF GraspNet: Variational Grasp Generation for Object Manipulation“ in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, S. 2901–2910, doi: 10.1109/ICCV.2019.00299.

- [17] Z. Dong *et al.*, „PPR-Net: Point-wise Pose Regression Network for Instance Segmentation and 6D Pose Estimation in Bin-picking Scenarios“ in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019, S. 1773–1780, doi: 10.1109/IROS40897.2019.8967895.
- [18] Y. Xiang, T. Schmidt, V. Narayanan und D. Fox, „PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes“, 1. Nov. 2017. [Online]. Verfügbar unter: <https://arxiv.org/pdf/1711.00199>.
- [19] X. Deng, Y. Xiang, A. Mousavian, C. Eppner, T. Bretl und D. Fox, „Self-supervised 6D Object Pose Estimation for Robot Manipulation“ in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 2020, S. 3665–3671, doi: 10.1109/ICRA40945.2020.9196714.
- [20] Q. Shao *et al.*, „Suction Grasp Region Prediction Using Self-supervised Learning for Object Picking in Dense Clutter“ in *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*, Singapore, 2019, S. 7–12, doi: 10.1109/ICMSR.2019.8835468.
- [21] K. Bousmalis *et al.*, „Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping: 21-25 May 2018“ in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, 2018.
- [22] M. Sundermeyer, A. Mousavian und D. F. R. Triebel, *Contact-GraspNet: Efficient 6-DoF Grasp Generation in Cluttered Scenes* (Zugriff am: 13. Dezember 2021).
- [23] D. Morrison, P. Corke, J. Leitner, „Learning robust, real-time, reactive robotic grasping“ (en), *The International Journal of Robotics Research*, Jg. 39, 2-3, S. 183–201, 2020, doi: 10.1177/0278364919859066.
- [24] M. Gualtieri und R. Platt, „Robotic Pick-and-Place With Uncertain Object Instance Segmentation and Shape Completion“ (eng), *IEEE robotics and automation letters*, Jg. 6, Nr. 2, S. 1753–1760, 2021, doi: 10.1109/lra.2021.3060669.
- [25] L. Ragginger, *Konzeptentwicklung einer Roboterzelle zur Entnahme und Bereitstellung von Bauteilen aus chaotischer Lagerung: Bachelorarbeit*. Köln, 2021.
- [26] VDI 2860:1990-05: *Montage- und Handhabungstechnik; Handhabungsfunktionen, Handhabungseinrichtungen; Begriffe, Definitionen, Symbole*, 2860, Berlin, Mai. 1990. [Online]. Verfügbar unter: <https://www.beuth.de/de/technische-regel/vdi-2860/847032?websource=vdin>
- [27] A. Wieland und C. Marcus Wallenburg, „Dealing with supply chain risks: Robustheit - Def.“, *Int Jnl Phys Dist & Log Manage*, Jg. 42, Nr. 10, S. 887–905, 2012, doi: 10.1108/09600031211281411.
- [28] G. Leao, *Robotic Bin Picking of Entangled Tubes*. Dissertation. Porto, Portugal, 2020.
- [29] K. Harada, N. Yamanobe, K. Nagata und T. Ueshiba, „Project on Development of a Robot System for Random Picking-Grasp/manipulation planner for a dual-arm manipulator“ in *2014 IEEE/SICE International Symposium on System Integration (SII 2014): Tokyo, Japan, 13 - 15 December 2014*, 2014, doi: 10.1109/sii.2014.7028104.
- [30] M. Gualtieri, A. Pas, K. Saenko und R. Platt, „High precision grasp pose detection in dense clutter“ in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, 2016, S. 598–605, doi: 10.1109/IROS.2016.7759114.
- [31] S. Song, A. Zeng, J. Lee und T. Funkhouser, „Grasping in the Wild: Learning 6DoF Closed-Loop Grasping From Low-Cost Demonstrations“, *IEEE Robot. Autom. Lett.*, Jg. 5, Nr. 3, S. 4978–4985, 2020, doi: 10.1109/LRA.2020.3004787.
- [32] S. Nördinger, *BMW macht's vor - So lohnt sich der ‚Griff in die Kiste‘*. [Online]. Verfügbar unter: <https://www.produktion.de/technik/so-lohnt-sich-der-griff-in-die-kiste-111.html> (Zugriff am: 11. Dezember 2021).
- [33] TensorFlow, *TensorFlow*. [Online]. Verfügbar unter: <https://www.tensorflow.org/> (Zugriff am: 10. Dezember 2021).
- [34] GitHub, *TensorFlow 2 Detection Model Zoo*. [Online]. Verfügbar unter: https://github.com/tensorflow/models/blob/c1ff413ad1cdd456b6da7fc213d9dcebc3957c5a/research/object_detection/g3doc/tf2_detection_zoo.md (Zugriff am: 1. Oktober 2021).
- [35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed und A. B. C.-Y. Fu, „SSD: Single Shot MultiBox Detector“, Jg. 9905, S. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.
- [36] T.-Y. Lin *et al.*, „Microsoft COCO: Common Objects in Context“, 1. Mai 2014. [Online]. Verfügbar unter: <https://arxiv.org/pdf/1405.0312>.