Einsatz von künstlichen neuronalen Netzen zur Verbesserung der Autonomisierung von Fahrerlosen Transportfahrzeugen

Edgar SCHERSTJANOI[1], Patrick BODEN[2],
Daniel GRÖLLICH[1], Ralf HUPFER[3], Martin SCHMAUDER[1]

[1] Professur für Arbeitswissenschaft, Institut für Technische Logistik und Arbeitssysteme, Fakultät Maschinenwesen, Technische Universität Dresden
Marschnerstraße 39, D-01307 Dresden
[2] Professur für Technische Logistik, Institut für Technische Logistik und Arbeitssysteme, Fakultät Maschinenwesen, Technische Universität Dresden
Münchner Platz 3, D-01187 Dresden
[3] GLOBALFOUNDRIES Dresden Module One LLC & Co. KG
Wilschdorfer Landstraße 101, D-01109 Dresden

Kurzfassung: Fahrerlose Transportsysteme [FTS] konnten sich in diversen Industriezweigen für den Transport von Gütern etablieren. Da sich FTS und Mensch einen gemeinsamen Arbeitsraum teilen, kommt es im industriellen Alltag häufig zu gegenseitigen Behinderungen. Um die Flexibilität und Robustheit von FTS zu erhöhen wird daher ein höherer Grad an Autonomie angestrebt. Die Fahrzeuge sollen in einem vorgegebenen Ausmaß selbstständig auf Umgebungseinflüsse reagieren können. Hierzu zählen das Ausweichverhalten bei Hindernissen oder die Lagebestimmung von Objekten (bspw. einem Ladungsträger). Zur Erfassung der Umgebung können Kameras und an den Anwendungsfall angepasste Bildverarbeitungsalgorithmen genutzt werden. Aufgrund der hohen Vielfalt an identifizierbaren Objekten und der zunehmenden Erkennungsqualität bietet sich der Einsatz von neuronalen Netzen an. Dabei hat die Auswahl der zu detektierenden Objekte wesentlichen Einfluss auf die Güte der Bilderkennung. Anhand einer exemplarischen Anwendung in der Halbleiterindustrie wird dieser Einsatz untersucht.

Schlüsselwörter: Objekterkennung, neuronale Netze,
fahrerlose Transportfahrzeuge, Bildverarbeitung

1. Motivation

Fahrerlose Transportsysteme [FTS] sind ein fester Bestandteil moderner Logistiksysteme. Da FTS und Mensch häufig einen gemeinsamen Arbeitsraum nutzen, kommt der Interaktion des Fahrerlosen Transportfahrzeugs [FTF] mit seiner Umgebung eine besondere Bedeutung zu. Informationen über den Umgebungszustand können vielfältig für die Steuerung der FTF berücksichtigt werden und bspw. ein an die jeweilige Situation angepasstes Ausweichverhalten bei Hindernissen ermöglichen. Der höhere Grad an Autonomie kann zu einer Verbesserung der Flexibilität von FTS beitragen.

Diverse Sensoren können für die Erfassung der Umgebung genutzt werden. So werden bereits bspw. Ultraschall- und Infrarotsensoren eingesetzt, um die Beschaffenheit der Fahrstrecke zu kontrollieren oder Hindernisse zu erkennen (Bostelmann

et al., 2005; Ray et al. 2008). Solche Sensoren werden verwendet, da ihre Verarbeitungsgeschwindigkeit, Genauigkeit und Aussagekraft den hohen Anforderungen entsprechen und sicherheitsrelevante Anwendungen normkonform implementiert werden können. Herkömmliche RGB-Kameras kamen bisher nur eingeschränkt zum Einsatz, weil der oftmals komplexe Informationsgehalt eines Bildes nur unter hoher Rechenleistung oder eher statischen Umgebungsbedingungen ausreichend genau interpretiert werden konnte.

Mittlerweile werden sogenannte Convolutional Neural Networks [CNN] genutzt, um ausgewählte Objekte in Kamerabildern zu erkennen. Sie können dabei unter deutlich variierenden Umständen wie Lichtverhältnissen, Perspektiven oder Bildrauschen eine hohe Erkennungsgenauigkeit erreichen. Dank stark weiterentwickelter Rechentechnik, wie zum Beispiel in Grafikkarten (Stirgl et al., 2010), sind bei Objektdetektionen nun Echtzeitanwendungen realisierbar. Konkurrierende Algorithmen überzeugen mit Ergebnissen basierend auf umfangreichen Datensätzen (Lin et al. 2014; Russakovsky et al. 2015) durch hohe eine Präzision und Trefferquote. Üblicherweise sind darin allgemeine Objektklassen definiert: Menschen, Pflanzen und Tiere, aber auch Alltagsgegenstände, Möbel oder Fahrzeuge. Da die Prinzipien zur Erkennung weitestgehend unabhängig zur betreffenden Klasse verstanden werden können, liegt der Gedanke nahe, auch für FTS relevante Gegenstände zu detektieren, um die Informationsverarbeitung zu erweitern.

2. Anwendungsumgebung

Als Anwendungsszenario wird das FTS einer Halbleiterfabrik betrachtet. Da die Wafer im Laufe der Produktion mehrere hundert Male zwischen Produktions- und Lagerplätzen transportiert werden müssen, ist die Effizienz logistischer Prozesse von besonderer Relevanz. Als Ladungsträger werden in modernen Halbleiterfabriken üblicherweise sogenannte Front Opening Unified Pods [FOUPs] verwendet. FOUPs bilden darum die erste der untersuchten Objektklassen. Abbildung 1 zeigt verschiedene Ansichten eines FOUPs.

Abbildung 1: *Verschiedene Ansichten eines typischen Transportbehälters von Wafern (Front Opening Unified Pod – FOUP).*

Die Ladungsträger werden unter anderem in Regalen gelagert, darum spielen Regale für den Transportprozess eine ebenso wichtige Rolle. Aufgrund der grafischen Unterschiede zwischen Regalen und FOUPs werden im Rahmen der Untersuchung Regale als zweite Objektklasse behandelt. In Abbildung 2 werden diese Unterschiedliche deutlich: Regale sind nicht auf Anhieb zu erkennen. Sie sind stets unterschiedlich gefüllt, der Hintergrund variiert und es besteht eine hohe Verwechslungsgefahr.

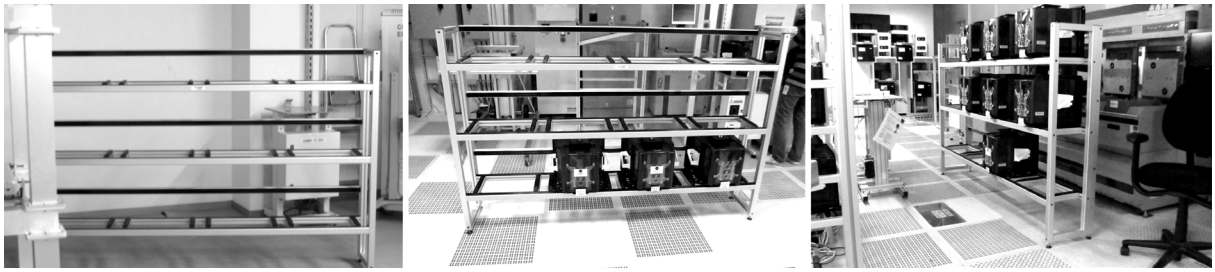

Abbildung 2: *Aufnahmen von Regalen, die zur Zwischenlagerung von FOUPs verwendet werden. Da Positionen und Füllstände stark variieren ist es schwer grafisch eindeutige Merkmale ableiten zu lassen.*

In Arbeitsumgebungen von Mensch und Maschine sind auch sehr spontane Änderungen des Arbeitsumfeldes möglich, die von FTF erkannt und richtig interpretiert werden müssen. Reparatur- oder Installationsarbeiten werden beispielsweise durch Beschilderung ausgewiesen. Da Schilder als ebene Objekte verstanden werden können, die in einem Kamerabild lediglich durch perspektivische Verzerrung projiziert werden, soll damit die dritte Objektklasse für die folgende Untersuchung gegeben sein. Abbildung 3 zeigt das beispielhaft verwendete Warnschild in unterschiedlichen Perspektiven.

Abbildung 3: *Zur Untersuchung verwendetes Hinweisschild, welches eingesetzt wird, wenn beispielsweise nach der Bodenreinigung eine Rutschgefahr besteht.*

Jede der drei Objektklassen FOUP [FOUP], Regal [SHELF] und Schild [SIGN] unterscheidet sich wesentlich, weswegen auch stark unterschiedliche Versuchsergebnisse erwartet werden. Durch die folgenden Versuche kann ein Ansatz zur Bestimmung der Eignung unterschiedlicher Objektklassen für eine kamerabasierte Detektion formuliert werden.

3. Versuchsaufbau

Die Arbeit von Redmon et al. (2016) erreicht im Vergleich aktueller Detektoren eine hohe durchschnittliche Genauigkeit bei vergleichsweise sehr hoher Geschwindigkeit in der Datenverarbeitung. Ebenso werden Quellcode und Framework zur Verfügung gestellt, weswegen eine Anwendung auf neu definierte Objektklassen gut umsetzbar ist. Das „You Only Look Once“ [YOLO] - Netz ist darin in zwei Varianten verfügbar. Neben der regulären Implementierung, kann eine kompaktere Version [YOLO-Tiny] verwendet werden, die eine höhere Verarbeitungsgeschwindigkeit, jedoch auch eine geringere positive Erkennungsrate aufweist.

Für die Trainingsphase wurde für jede Klasse ein ~300 Bilder großer Trainingsdatensatz erstellt. Zur Bestimmung der Genauigkeit diente ein dazu unabhängiger

Testdatensatz, bestehend aus zufällig gewählten Auszügen aus Videosequenzen, in denen die betreffenden Objekte enthalten waren.

Untersuchungen zur Klasse der FOUPs umfassen 3 verschiedene Testdatensätze: (1) Bilder mit ausschließlich einzelnen FOUPs, die vollständig sichtbar sind [SNG], (2) Bilder mit einzelnen FOUPs, die bis zu 50% von anderen Objekten verdeckt sind [OCC] und (3) Bilder mit 2 FOUPs, die sich manchmal gegenseitig verdecken, aber jeweils zu mindestens 50% sichtbar sind [DBL]. Zur Evaluation der Detektion von Regal und Warnschildern werden Testdatensätze verwendet, in denen pro Bild eine Instanz vorhanden ist. Für jeden Testdatensatz wurden 100 Bilder zusammengeführt.

Zur Auswertung werden für jeden Datensatz alle richtig [TP] und falsch markierten [FP] Instanzen gezählt. TP werden in das Verhältnis zum Ideal - dem Erkennen aller geforderten Instanzen - prozentual angegeben. Da FP auch mehrfach pro Bild vorkommen können soll die absolute Anzahl den nötigen Aufschluss zur Qualität geben. Darüber hinaus wurde der Testdatensatz zu den drei Klassen jeweils mit einem 200 Bilder großen Datensatz ergänzt, der keine der gesuchten Instanzen enthält, weswegen sich die Anzahl der FP auf insgesamt 300 Bilder bezieht.

Jeder Versuch wurde einerseits mit YOLO, andererseits mit YOLO-Tiny durchgeführt, wodurch qualitative aber auch systematische Unterschiede verdeutlicht werden können.

4. Ergebnisse

Am Beispiel der FOUP-Klasse sind in Tabelle 1 die Ergebnisse gegenübergestellt. Da auch Aussagen zur Vertrauenswürdigkeit zu jeder Detektion produziert werden, ist jeweils die Ergebnismenge abgebildet, sollte ein Detektionsschwellwert über 50%, 70% bzw. 90% Konfidenz angenommen werden. Je niedriger der Schwellwert, desto höher fällt die Präzision aus, bei steigender Anzahl von falsch-positiven Markierungen [FP].

Tabelle 1: *Ergebnisse bei zur Auswertung der Testfälle zur Erkennung von FOUPs in Kamerabildern mit YOLO und YOLO-Tiny unter Verwendung von unterschiedlichen Schwellwerten. SNG umfasst ausschließlich einzelne, vollständig sichtbare Instanzen. OCC enthält Bilder, in denen einzelne FOUPs nur verdeckt vorkommen, wohingegen mit DBL doppelt vorkommende FOUPs in Bildern untersucht wurden.*

	FOUP SNG		**FOUP OCC**		**FOUP DBL**	
	TP (%)	FP (#)	TP (%)	FP (#)	TP (%)	FP (#)
YOLO						
.5	100	10	94	17	81	30
.7	96	4	84	8	72	0
.9	24	0	12	0	47	0
YOLO Tiny						
.5	72	8	58	7	50	12
.7	69	5	47	5	40	8
.9	61	1	34	2	32	0

Die Erkennung einzelner FOUPs [SNG] erzeugt gute Ergebnisse, bei niedrigem Schwellwert durch hohe Vollständigkeit: Alle gesuchten Instanzen wurden erfolgreich markiert. Wird der Schwellwert leicht erhöht sinkt zwar erwartungsgemäß die Treffer-

quote jedoch kann auch der Anteil von falsch-positiven Detektionen halbiert werden. Auffällig ist jedoch die stark fallende Anzahl von TP bei sehr hohem Schwellwert. Diese Eigenschaft lässt sich mit YOLO-Tiny nicht produzieren. Die Ergebnisse dort sind stets mit hohen Werten zur Vertrauenswürdigkeit hinterlegt und erreichen auch bei einem Schwellwert > 90% eine Trefferquote von 61%. In Abbildung 4 sind beispielhaft erfolgreich erkannte Bilder zu den jeweiligen Testfällen dargestellt.

Abbildung 4: *Detektionen von FOUPs in exemplarischen Auszügen von Videoaufnahmen zu den jeweiligen Szenarien SNG (li.), OCC (mi.) und DBL (re.). Die roten Markierungen umfassen den detektierten Bildbereich. Innerhalb der Markierung ist der dazugehörige Konfidenzwert abgebildet.*

Die Experimente mit den Objektklassen Schild und Regal sollen einen weiteren wichtigen Aspekt beim Einsatz von Objekterkennung mit CNN verdeutlichen: die Objektbeschaffenheit. Schilder sind ebene Objekte auf meist farblich auffällig gekennzeichneten Flächen, ohne geometrische oder materielle Besonderheiten, Regale wiederum sind unterschiedlich gefüllt, haben je nach Positionierung einen variierenden Hintergrund, bestehen aus einem Material, was oft für weitere Objekte genutzt wird und haben innerhalb ihrer konvexen Hülle im Bild einen überdurchschnittlich geringen Bildanteil. Schilder sind wesentlich deutlicher erkennbar, Regale fallen im Bild hingegen wenig auf. Diese eher menschliche Interpretation lässt sich auch in den Ergebnissen unter Anwendung von CNN ablesen. In Tabelle 2 sind die erhobenen Resultate zusammengefasst.

Tabelle 2: *Ergebnisse bei der Objektdetektion von Schildern [SIGN] und Regalen [SHELF] mit den beiden CNNs YOLO und YOLO-Tiny unter Verwendung unterschiedlicher Schwellwerte.*

	SIGN		**SHELF**	
	TP (%)	FP (#)	TP (%)	FP (#)
YOLO				
.5	96	2	23	9
.7	85	0	15	4
.9	02	0	00	0
YOLO Tiny				
.5	78	0	04	0
.7	73	0	01	0
.9	65	0	00	0

Schilder [SIGN] lassen sich sehr gut erkennen, bei (verglichen mit FOUPs) deutlich weniger falsch-positiven Detektionen. Auch YOLO-Tiny erreicht Ergebnisse, die im Rahmen einer Anwendung vielversprechend sind. Regale [SHELF] hingegen konnten, selbst bei geringem Schwellwert nur in jedem vierten Fall der YOLO An-

wendung erfolgreich erkannt werden. YOLO-Tiny scheint unter den gegebenen Umständen dafür deutlich unbrauchbar.

5. Diskussion

Die Ergebnisse zeigen, dass die Erkennungsrate in Abhängigkeit der Objektklasse stark variiert. Somit kommen manche Objekte vorerst nicht für eine Detektion in Frage. Für den Einsatz in einem industriellen Umfeld muss die Qualität der Detektion insgesamt verbessert werden. Eine Fehldetektion könnte zu unvorhergesehenen Aktionen der FTF führen, welche im Zweifelsfall Menschen verletzen oder Objekte beschädigen könnten.

Um die Ergebnisse zu verbessern könnte sowohl die Parametrierung des neuronalen Netzes als auch seine Struktur angepasst werden. Da die Parameter des Netzes noch nicht angepasst und ausreichend untersucht wurden, ist absehbar, dass damit auch weitere qualitative Verbesserungen einhergehen. Ähnliches gilt auch für die Größe der Trainingsdatenmenge und Trainingsdauer, welche vergleichsweise gering gewählt wurden.

Die Entwickler von YOLO fokussieren sich bisher auf sehr weitgefasste Objektklassen. Da relevante Objekte im Anwendungsumfeld jedoch stets stark ähnelnde Formen aufweisen, wäre eine angepasste Entwicklung eines entsprechenden neuronalen Netzes naheliegend.

6. Literatur

Bostelman RV, Hong TH, & Madhavan R (2005) Towards AGV safety and navigation advancement obstacle detection using a TOF range camera. In Advanced Robotics, 2005. ICAR'05. Proceedings., 12th International Conference on. IEEE. 460-467.

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, & Zitnick CL (2014) Microsoft coco: Common objects in context. In European conference on computer vision. Springer, Cham. 740-755.

Ray, A. K., Gupta, M., Behera, L., & Jamshidi, M (2008) Sonar based autonomous automatic guided vehicle (AGV) navigation. In System of Systems Engineering, 2008. SoSE'08. IEEE International Conference on. IEEE. 1-6.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A (2016) You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE. 779-788.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC and Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. IJCV.

Strigl, D., Kofler, K., & Podlipnig, S (2010) Performance and scalability of GPU-based convolutional neural networks. In Parallel, Distributed and Network-Based Processing (PDP), 2010 18th Euromicro International Conference on. IEEE. 317-324.

Förderhinweis

Die Arbeit wurde im Rahmen des Forschungsprojekts „Erforschung von Grundlagen und Konzepten zur Gestaltung einer automatisch auf sich ändernde Anforderungen, hinsichtlich Produktionsvolumen und Produktmix, reagierende Halbleiterfabrik“ (Responsive Fab) durchgeführt und über die Sächsische Aufbau-Bank aus Mitteln des Europäischen Fonds für regionale Entwicklung (EFRE) und des Freistaates Sachsen gefördert.

Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit interdisziplinär
analysieren – bewerten – gestalten

65. Kongress der
Gesellschaft für Arbeitswissenschaft

Professur Arbeitswissenschaft
Institut für Technische Logistik und Arbeitssysteme
Technische Universität Dresden

Institut für Arbeit und Gesundheit
Deutsche Gesetzliche Unfallversicherung

27. Februar – 1. März 2019

GfA-Press

Bericht zum 65. Arbeitswissenschaftlichen Kongress vom 27. Februar – 1. März 2019

Professur Arbeitswissenschaft, Institut für Technische Logistik und Arbeitssysteme, Technische Universität Dresden;
Institut für Arbeit und Gesundheit, Deutsche Gesetzliche Unfallversicherung, Dresden

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2019
ISBN 978-3-936804-25-6

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© GfA-Press, Dortmund
Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Konferenzband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Konferenzband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

Screen design und Umsetzung
© 2019 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de