Bei diesem Beitrag handelt es sich um einen wissenschaftlich begutachteten und freigegebenen Fachaufsatz („reviewed paper“).

doi.org/10.37544/1436-4980-2025-11-12-63
Datum der Einreichung: 15.09.2025
Datum der Annahme: 10.11.2025
Datum der Veröffentlichung: 17.12.2025

Computer Vision-gestützte Montageanleitung: systematisches Design

YOLO-unterstützte Echtzeit-Montageassistenz

J. Liang, A. Moriz, A. Göppert, R. H. Schmitt

ZUSAMMENFASSUNG Die Arbeit präsentiert das systematische Design eines Computer Vision-gestützten Montageanleitungssystems, das YOLO-basierte Objekterkennung mit metadatenbasierten Anleitungsmodellen integriert. Der Lösungsansatz liefert Mitarbeitern kontextspezifische Arbeitsanleitungen in Echtzeit, reduziert die kognitive Belastung und verbessert gleichzeitig die Genauigkeit der Aufgabenausführung. Das Konzept bildet die Grundlage für adaptive Assistenzsysteme in dynamischen Montageumgebungen.

STICHWÖRTER

Cognitive Engineering, Digitalisierung, Montage

Computer Vision-aided Assembly Instruction: Systematic Design

ABSTRACT This paper presents a systematic design of a computer vision-aided assembly instruction system that integrates YOLO-based object recognition with metadata-based instruction models. The approach delivers real-time and context-specific work instruction to workers, reducing cognitive workload while improving task execution accuracy. The concept establishes a foundation for adaptive assistance systems in dynamic assembly environments.

1 Ausgangssituation

Trotz des heutigen hohen Automatisierungsgrades in Fertigungsprozessen werden zahlreiche Aufgaben, darunter Wartungs-, Reparatur-, Montage- oder Einrichtungsarbeiten, manuell ausgeführt [1]. Unter diesen Tätigkeiten sind insbesondere Montageprozesse in hohem Maße von der Fähigkeit der Mitarbeiter abhängig, komplexe Anleitungen genau zu interpretieren und auszuführen. Eine zentrale Herausforderung ergibt sich aus der kognitiven Belastung, die durch herkömmliche textbasierte oder schematische Arbeitsanleitungen entsteht [2]. Die Mitarbeiter müssen kontinuierlich relevante Details aus langen und oft allgemeinen Beschreibungen herausfiltern, die Anleitungen mental auf die physische Arbeitsumgebung übertragen und sie während der Ausführung präziser Montagevorgänge im Gedächtnis behalten. Dies führt zu einer hohen kognitiven Belastung, die durch Informationsüberflutung, Gedächtnisbelastung und häufiges Wechseln zwischen Lesen und Handeln gekennzeichnet ist.

Die Herausforderung wird in Umgebungen mit hoher Produktvariabilität und steigenden Anforderungen an die Individualisierung noch verstärkt, wo die Anleitungen immer komplexer werden und sich in mehrere Varianten verzweigen. Daher sind Montagearbeiter einem erhöhten Risiko von Fehlinterpretationen und Fehlern, verminderter Effizienz und langsameren Lernkurven ausgesetzt [3]. Diese Probleme verdeutlichen eine grundlegende Einschränkung traditioneller Anleitungsformate: Diese Formate bieten zwar eine umfassende Dokumentation, unterstützen jedoch nicht den Bedarf der Mitarbeiter an kontextspezifischen, mühelosen Anleitungen während der Ausführung der Aufgaben [4]. Im Allgemeinen verwenden viele produzierende Unternehmen immer noch Dokumente, die auf dem traditionellen Format basieren. Auch wenn einige Unternehmen behaupten, dass die Bereitstellung von Arbeitsanleitungen in digitaler PDF-Form ebenfalls als Digitalisierung betrachtet werden kann, ist das PDF-Format nach wie vor für den Ausdruck oder die Anzeige auf dem Bildschirm eines elektronischen Geräts konzipiert. Dieser Art der „Digitalisierung” mangelt es an Flexibilität und Anpassungsfähigkeit. Wenn sich einzelne Komponenten in den Anleitungen ändern, beispielsweise wenn ein Bild ersetzt oder einzelne Parameter angepasst werden müssen, muss das gesamte Dokument durch eine aktualisierte Version ersetzt werden. Neben dem unverhältnismäßig hohen Zeitaufwand erfordert die Erstellung einer aktualisierten Version auch Ressourcen und Fachwissen von erfahrenen Mitarbeitern, um das Know-how in den Anleitungen darzustellen.

Um die Qualität, Genauigkeit und Effizienz der Aufgaben in modernen Montagesystemen zu gewährleisten, ist es daher notwendig, die mit der Verwendung von Anleitungen verbundene kognitive Arbeitsbelastung zu reduzieren. Es sind Forschungsarbeiten erforderlich, um systematische Ansätze zu identifizieren, die den Mitarbeitern die richtigen Informationen zur richtigen Zeit und im richtigen Format zur Verfügung stellen, wodurch der mentale Aufwand verringert, und eine nachhaltig hohe Leistungsqualität ermöglicht wird.

2 Zielsetzung

Die Zielsetzung dieser Arbeit ist es, ein Computer Vision-gestütztes Montageanleitungssystem zu entwickeln und zu evaluieren, das die kognitive Arbeitsbelastung bei der manuellen Mon-

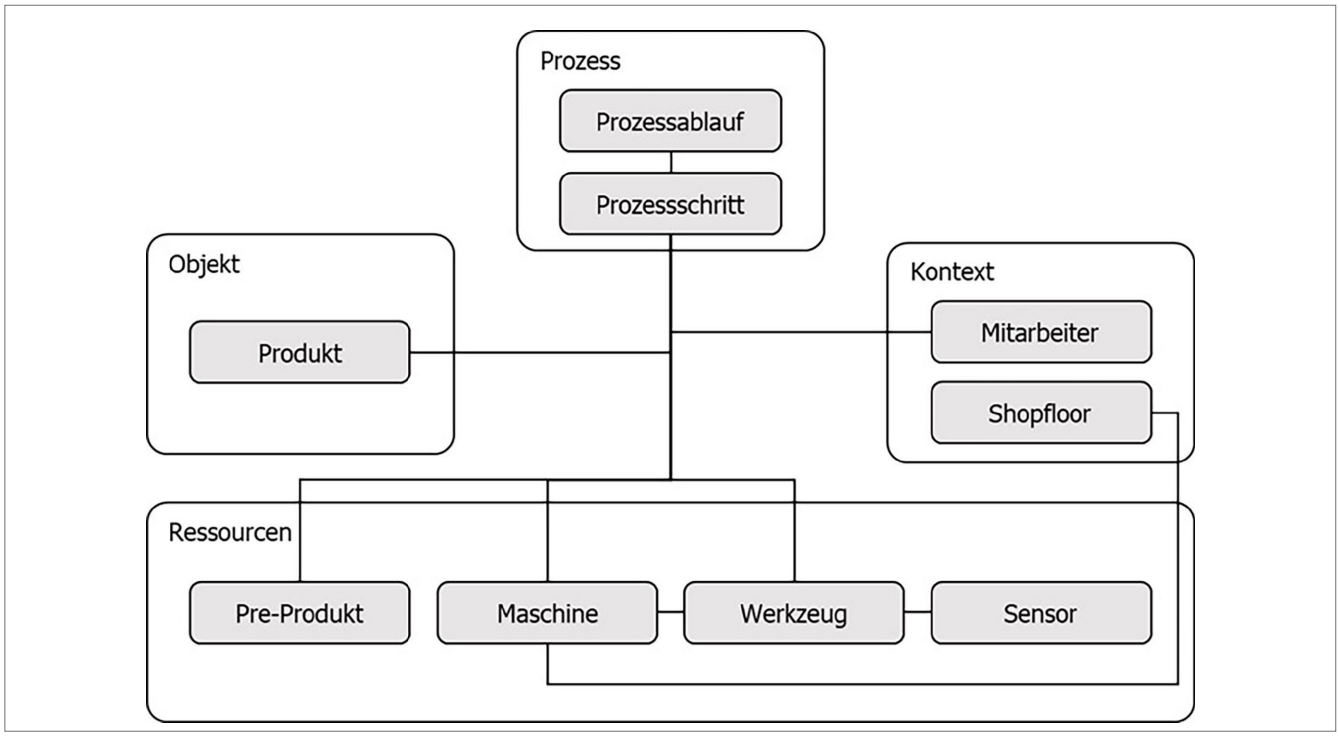

Bild 1 Relevante Datenklassen und Zuordnung in die Kategorien. *Grafik: [20, 21]*

tage reduziert, indem es den Mitarbeitern kontextspezifische und genaue Anleitungen gibt. Das Anleitungssystem soll in die Lage sein, dieselbe Arbeitsumgebung wie der Mitarbeiter wahrzunehmen, indem es mithilfe von Computer Vision Bauteile, Werkzeuge und Montagezustände in Echtzeit erkennt. Durch die Überbrückung der Lücke zwischen Erkennung und Anleitung kann das System automatisch den relevanten Prozessschritt bestimmen und dem Mitarbeiter eine genau auf die aktuelle Situation zugeschnittene Anleitung geben.

Das gezielte System befasst sich mit zwei zentralen Herausforderungen – kognitive Arbeitsbelastung und dynamische Montagebedingungen. Das erste Ziel des Systems besteht darin, die kognitive Arbeitsbelastung zu reduzieren, indem es den Mitarbeitern die Notwendigkeit nimmt, Montageanleitungen manuell zu durchsuchen, anwendbare Varianten zu identifizieren und wichtige Details während der Ausführung der Aufgabe im Gedächtnis zu behalten. Stattdessen werden die erforderlichen Informationen in einem kontextbezogenen und sofort zugänglichen Format dargestellt, das direkt auf die Teile und Aktionen abgestimmt ist, die im Sichtfeld des Mitarbeiters sichtbar sind. Zweitens zielt das System darauf ab, die Genauigkeit und Effizienz der Aufgaben zu gewährleisten, indem es sicherstellt, dass die Anleitungen nicht nur genau sind, sondern sich auch dynamisch an unterschiedliche Montagebedingungen anpassen. Dies ist besonders relevant in Umgebungen, die durch Produktvielfalt, individuelle Anpassungen und häufige Prozessänderungen gekennzeichnet sind, in denen statische Anleitungen oft keine ausreichende Unterstützung bieten.

Um diese Ziele zu erreichen, werden in diesem Beitrag strukturierte Anleitungsmodelle verwendet, die Montageschritte, Teile, Werkzeuge und Prozessspezifikationen explizit miteinander verknüpfen. Diese Modelle bilden die Grundlage des adaptiven Anleitsystems und ermöglichen es, die Ergebnisse der Computer Vision-Erkennung systematisch den entsprechenden Anleitungen zuzuordnen. Die Integration von Erkennung und Anleitung schafft somit einen Regelkreis, in dem das System seine Ausgabe kontinuierlich an den Echtzeitkontext des Mitarbeiters anpasst. Dadurch werden die Anleitungen nicht nur konkreter, sondern auch intuitiver, was den mentalen Aufwand verringert, der erforderlich ist, um die Lücke zwischen abstrakten Informationen und spezifische Handlungen zu schließen.

Im Allgemeinen wird eine systematische Grundlage für eine adaptive und Mitarbeiterorientierte Anleitungserteilung bei der manuellen Montage angestrebt. Durch die Einbettung von Kontextbewusstsein in Anleitungssysteme leistet dieser Arbeit einen Beitrag zu einer neuen Generation von Assistenztechnologien, die direkt auf die kognitiven Herausforderungen menschlicher Mitarbeiter eingehen. Solche Systeme können die Ausführung von Aufgaben in Bezug auf Geschwindigkeit und Genauigkeit erheblich verbessern und gleichzeitig zum Wohlbefinden der Mitarbeiter beitragen, indem sie unnötige kognitive Belastungen verringern.

3 Stand der Technik

In diesem Abschnitt wird die aktuelle Entwicklung der relevanten Technologien diskutiert, wobei der Schwerpunkt auf Computer Vision (cf. Abschnitt 3.1) und Datenmodellierung für kontextspezifische Arbeitsanleitung (cf. Abschnitt 3.2) liegt. Durch die Analyse des Stands der Technik können der Forschungsbedarf ermittelt und die Entwicklungsrichtung festgelegt werden (cf. Abschnitt 3.3).

3.1 Computer Vision in der Produktion

In den letzten Jahren hat sich Computer Vision (CV) zu einer zentralen Technologie in der Produktion entwickelt. CV-Systeme basieren auf den Bereichen Bildverarbeitung und maschinelles Lernen und ermöglichen es Produktionssystemen, visuelle Infor-

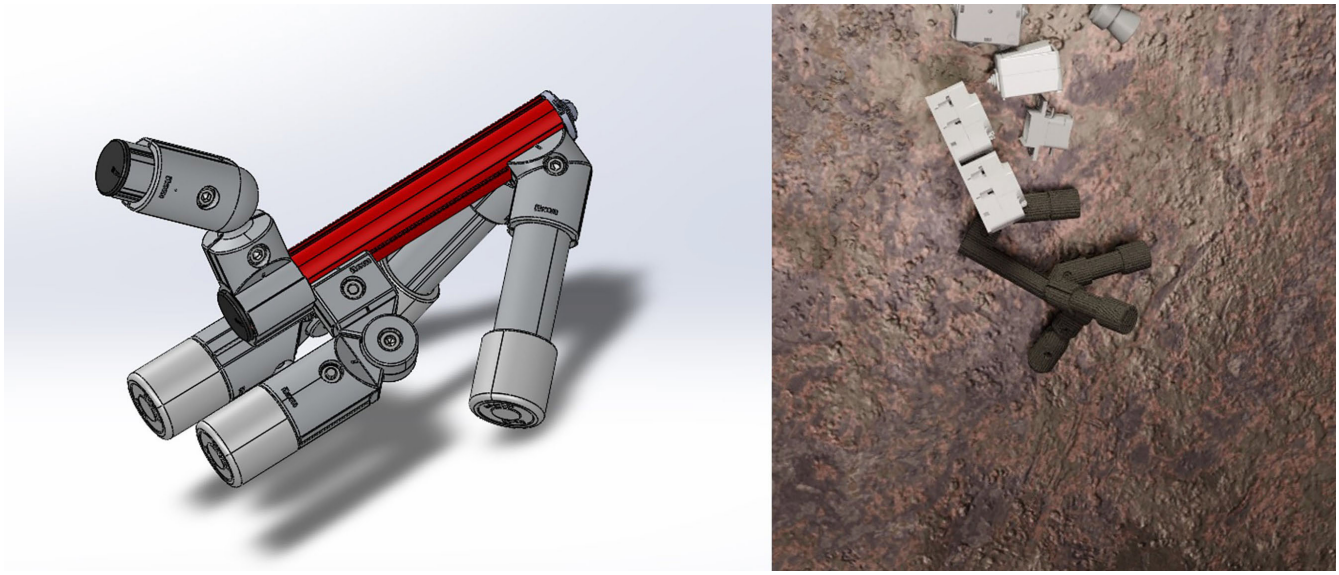

Bild 2 CAD Modell von der Baugruppe und die synthetischen Bilder zum Algorithmus-Training. *Grafik: RWTH Aachen*

mationen von optischen Sensoren (i.d.R. Kameras) zu interpretieren und aussagekräftige Erkenntnisse für die Entscheidungsfindung und Prozesssteuerung zu gewinnen [5]. Im Produktionsprozess wird CV eingesetzt, um die Qualitätssicherung zu automatisieren und die Mitarbeiter, die bestimmten Produktionsprozesse durchführen, intelligent zu unterstützen.

Die Anwendung von CV in der Produktion sind vielfältig. Einer der etabliertesten Anwendungsfälle ist die visuelle Qualitätskontrolle, bei der Kamerasysteme automatisch optische Fehler wie Kratzer, Fehlausrichtungen oder fehlende Komponenten mit einer höheren Zuverlässigkeit und Konsistenz als manuelle Kontrollen erkennen [6]. Ein weiterer wichtiger Bereich ist die Prozessüberwachung, bei der visuelle Daten verwendet werden, um Materialflüsse, Maschinenzustände oder Montagesequenzen in Echtzeit zu verfolgen [7]. Darüber hinaus unterstützt CV menschenzentrierte Aufgaben, beispielsweise durch die Überwachung der Ergonomie, die Unterstützung bei der Mitarbeiterschulung oder die Ermöglichung adaptiver Anleitungssysteme, die auf die aktuellen Tätigkeiten der Mitarbeiter reagieren.

Die zunehmende Komplexität und Varianten von Produkten haben die Bedeutung von CV weiter erhöht [8]. Klassische sensorbasierte Ansätze für Qualitätskontrolle (beispielsweise Lichtschranken, Ultraschallsensor, 3D-Scanner) sind oft auf vordefinierte Messaufgaben beschränkt und bieten keine Flexibilität im Umgang mit unvorhergesehenen Abweichungen [9]. Im Gegensatz dazu bieten moderne CV-Systeme – insbesondere solche, die auf Deep Learning basieren – die Möglichkeit, eine Vielzahl von Szenarien zu verallgemeinern, wodurch sie sich besonders für dynamische Produktionsumgebungen eignen [8]. Diese Anpassungsfähigkeit ermöglicht die Erkennung verschiedener Teile, Werkzeuge und Montagezustände und schafft neue Möglichkeiten für adaptive und kontextspezifische Unterstützungssysteme für Mitarbeiter.

Mit Deep Learning und insbesondere der Vorstellung von Convolutional Neural Networks (CNNs) hat sich die visuelle Objekterkennung in der Produktion rasant weiterentwickelt. Diese Modelle sind in der Lage, hierarchische Merkmalsdarstellungen automatisch aus Daten zu lernen, was eine robuste Erkennung in komplexen Umgebungen, zum Beispiel unterschiedliche Lichtverhältnisse, ungleichmäßiger Hintergrund oder verschiedene Positionierungen des Objekts, ermöglicht [10].

Unter den verschiedenen Deep-Learning-basierten Ansätzen zur Objekterkennung hat der Algorithmus YOLO in produktionsorientierten Anwendungen Wesentlichkeit erlangt [11]. YOLO steht für „You Only Look Once" (Man schaut nur einmal hin). Es handelt sich um ein Modell zur Objekterkennung und Bildsegmentierung, das von Redmon et al. an der University of Washington entwickelt wurde. Der Vorteil von YOLO liegt in seiner Fähigkeit, Objekte in Echtzeit mit hoher Genauigkeit zu erkennen, wodurch es sich besonders gut für schnelllebige Montageumgebungen eignet, in denen das System sofort und zuverlässig reagieren muss [12]. Im Gegensatz zu anderen Methoden wie Region-Proposal-Network (RPN) behandelt YOLO die Objekterkennung als ein einziges Regressionsproblem und sagt in einem Durchlauf durch das Netzwerk direkt Begrenzungsrahmen und Klassenwahrscheinlichkeiten voraus. Dieses Design führt zu erheblichen Geschwindigkeitsvorteilen bei gleichzeitig wettbewerbsfähiger Erkennungsleistung [13].

Die Kernfunktion von YOLO besteht darin, das Eingabebild in ein Raster zu unterteilen, wobei jede Rasterzelle für die Vorhersage von Begrenzungsrahmen und entsprechenden Wahrscheinlichkeiten für Objekte zuständig ist, deren Mittelpunkte innerhalb der Zelle liegen. Durch diesen einheitlichen Ansatz entfällt die Notwendigkeit einer separaten Region-Proposal-Phase, die RPN benötigt, so dass die Erkennung in einem einzigen Durchlauf durch das neuronale Netzwerk erfolgen kann [11].

Die YOLO-Modelle verwenden CNNs als Basis, um hierarchische Merkmalskarten aus dem Eingabebild zu extrahieren. Diese Merkmale werden anschließend durch den Algorithmus verarbeitet, die vorhersagen [11]:

- Koordinaten der Begrenzungsrahmen (x- und y-Achse, Breite, Höhe) zur Lokalisierung des Objekts
- Objectness Score, der die Wahrscheinlichkeit angibt, dass der Rahmen tatsächlich ein Objekt enthält
- Wahrscheinlichkeiten, die bestimmen, zu welcher Kategorie der Rahmen gehört

Der Trainingsprozess des YOLO-Algorithmus optimiert eine mehrteilige Verlustfunktion, die Lokalisierungsgenauigkeit, Konfi-

denzschätzung und Segmentierungsleistung ausbalanciert [11]. Anders als RPN behandelt YOLO die Objekterkennung als End-to-End-Optimierungsproblem, so dass das Modell gemeinsame Repräsentationen von Aussehen und räumlichen Informationen erlernen kann. Diese Struktur steigert nicht nur die Geschwindigkeit, sondern reduziert auch Fehlsegmentierung durch Hintergrundobjekte, da das Modell den globalen Kontext des gesamten Bildes berücksichtigt.

Im Bereich Produktion, insbesondere Montage, wird YOLO in mehreren Anwendungsfällen eingesetzt [14]:

- Teileverifikation: Erkennung spezifischer Bauteile während der Montage, um sicherzustellen, dass die richtige Variante verbaut wird.
- Werkzeugerkennung: Überprüfung, ob das passende Werkzeug verfügbar und korrekt positioniert ist, bevor ein Arbeitsschritt (zum Beispiel Schraubenanzug) ausgeführt wird.
- Mensch-Roboter-Kollaboration: Detektion von Mitarbeiterhänden oder Gesten, um eine sichere und adaptive Zusammenarbeit mit kollaborativen Robotern zu ermöglichen.

In dieser Arbeit wird YOLO angewendet, um die Funktion der Erkennung von bearbeiteten Bauteilen durch den Mitarbeiter zu ermöglichen. In Abschnitt 4.1 werden Einzelheiten dazu vorgestellt, wie der YOLO-Algorithmus trainiert wird und welche Parameter oder Informationen an das Montageanleitungssystem übertragen werden, um adaptive und kontextspezifische Arbeitsanleitungen zu ermöglichen.

3.2 Metadatenmodellierung für kontextspezifische Montageanleitung

Kontextspezifische Montageanleitungen zielen darauf ab, nur die Anleitungsinformation zu liefern, die für die Mitarbeiter zu der bestimmten Montageaufgabe relevant sind, wodurch der kognitive Aufwand für die Identifizierung und Interpretation wichtiger Details minimiert wird [15]. Um konkrete und exakte Information zu liefern, müssen solche Anleitungen auf Komponenten, Prozessen und persönliche Merkmale der Mitarbeiter (beispielsweise Qualifikation, Sprache, Expertise) angepasst sein, damit die bereitgestellten Anleitungen die kognitive Unterstützung verbessern. Während die Bedeutung der Kontextualisierung in der Fertigungsforschung in den letzten zehn Jahren betont wurde, bleibt eine konsistente Strukturierung relevanter Informationen nach wie vor unerlässlich. Frühere Forschungen definierten den Kontext anhand von benutzer-, umgebungs- und systembezogenen Attributen und erweiterten diese Definition anschließend, um die Prozessrelevanz hervorzuheben [16, 17, 18]. Aus diesen Forschungen hat sich ein gemeinsames Verständnis von Kontextinformationen herauskristallisiert, das sich in vier Kategorien zusammenfassen lässt: Objekt, Prozess, Ressource und Kontext.

Nuy et al. haben ein detailliertes Datenmodell für kontextspezifische Arbeitsanleitungen vorgestellt, das sich am Konzept des digitalen Zwillings orientiert [19]. Ihr prozessorientiertes Modell identifizierte vier wichtige Kategorien – Produkt, Ressource, Prozess und Kontext – die zusammen die Grundlage für die Bereitstellung von Anleitungen bilden. Dies steht in engem Einklang mit der umfassenderen Konzeptualisierung von Kontextinformationen im Fertigungsbereich. Die Definition jeder Kategorie lautet wie folgt:

- Objektdaten erfassen die Eigenschaften des Produkts, einschließlich Baugruppen, Unterbaugruppen und deren zugehörigen Bauteilen, sowie entsprechende geometrische Informationen.
- Prozessdaten definieren die für die Ausführung erforderlichen Schritte, Abläufe und aufgabenspezifischen Parameter.
- Ressourcendaten umfassen die für Montageaufgaben erforderlichen Werkzeuge, Maschinen und Hilfsstoffe.
- Kontextdaten repräsentieren mitarbeiterbezogene Attribute (beispielsweise Fähigkeiten und Erfahrungen), Umweltfaktoren (beispielsweise Lärm, Beleuchtung, Klima) und digitale Infrastruktur (beispielsweise Software und Netzwerkbedingungen).

Während die Theorie von Nuy et al. die Grundlage für die Zuordnung verschiedener Datenklassen während Produktionsprozessen liefert, haben Cramer et al. [20] und Liang et al. [21] das Meta-Modell für Produktionsdaten (MMPD) eingeführt, das relevante Datenklassen zur Strukturierung heterogener Produktionsdaten festlegt. Das MMPD bietet standardisierte Klassen für Produkt, Pre-Produkt, Prozessablauf, Prozessschritt, Maschine, Werkzeuge, Sensor, Mitarbeiter und Shopfloor und ermöglicht so eine domänenübergreifende konsistente Darstellung von Fertigungsdaten. Durch die Zuordnung der MMPD-Klassen zu den vier Kategorien Objekt, Prozess, Ressource und Kontext entsteht eine einheitliche Struktur, die Produktionsdaten mit den Informationsanforderungen von Anleitungssystemen verknüpft. Beispielsweise entsprechen die Klassen „Produkt“ und „Vorprodukt“ den Objektdaten, „Prozessschritt“ und „Prozessablauf“ den Prozessdaten, „Maschine“, „Werkzeug“ und „Sensor“ den Ressourcendaten und „Mitarbeiter“ und „Fertigungsumgebung“ den wichtigsten Aspekten der Kontextdaten.

Die Kombination der von den oben genannten Forschern vorgeschlagenen Ansätze ermöglicht es, kontextspezifische Arbeitsanleitungen systematisch durch Produktionsdatenmodelle zu unterstützen. Das Konzept des digitalen Zwillings, mit dem physikalische und funktionale Eigenschaften von Produkten, Prozessen und Ressourcen nachgebildet werden können, erleichtert den bidirektionalen Datenaustausch in Echtzeit und stellt sicher, dass Anleitungen anpassungsfähig und genau bleiben. Die Kombination von Kategorisierungsschemata mit einer Metadatenbasis wie dem MMPD harmonisiert nicht nur verschiedene Forschungsstränge, sondern bietet auch eine skalierbare und erweiterbare Grundlage für intelligente kognitive Unterstützung.

3.3 Entwicklungsbedarf

Wie im Abschnitt 2 dargelegt, besteht das Ziel dieser Arbeit darin, die Lücke zwischen der Erkennung der physischen Bauteile und der Bereitstellung kontextspezifischer Arbeitsanleitungen zu schließen. Die Kategorisierung kontextspezifischer Anleitungen bietet zwar einen konzeptionellen Rahmen für die Strukturierung von Anleitungsinformationen in die Bereiche Objekt, Prozess, Ressource und Kontext, doch gewährleisten diese Kategorien allein nicht, dass die richtige Anleitung zu der entsprechenden Situation und dem passenden Zeitpunkt dargestellt wird. Ebenso bietet MMPD eine standardisierte und umfassende Struktur für die Organisation heterogener Produktionsdaten, aber es bestimmt nicht von sich aus, welche Bauteile dieser Daten für den Mitarbeiter in einem Montageschritt relevant sind.

Die YOLO-basierte Objekterkennung spielt eine zentrale Rolle als Enabler, um diese Lücke zu schließen. Indem es dem System ermöglicht, der Bauteil durch Echtzeit-Erkennung von Baugruppen, Unterbaugruppen und Teilen zu „sehen“, stellt YOLO die

entscheidende Verbindung zwischen der physischen Arbeitsumgebung und dem digitalen Anleitungssystem her. Erkannte Bauteile können direkt auf die Objektdatenkategorie kontextspezifischer Anleitung abgebildet werden, die dann als Schnittstelle für die Identifizierung der relevanten Prozessdaten, zugehörigen Ressourcen und Kontextfaktoren dienen.

Eine systematische Lösung zur Zuordnung der Funktion der YOLO-basierten Objekterkennung und Zuordnung zu digitalen Anleitungen ist erforderlich. Dementsprechend zielt diese systematische Lösung darauf ab, die folgenden Funktionen bereitzustellen:

1. Automatisierung der Zuordnung physischer Bauteile zu digitalen Anleitungsdaten, um den manuellen Such- und Interpretationsaufwand für Anleitung zu reduzieren
2. Ermöglichung einer adaptiven Echtzeit-Anleitung, da das System seine Anleitungsdarstellung kontinuierlich entsprechend den erkannten Objekten und Prozesszuständen aktualisiert
3. Einsatz von MMPD, um sicherzustellen, dass Erkennungsergebnisse universell in strukturierte, prozessrelevante Informationen über verschiedene Produktionsumgebungen hinweg übertragen werden können

Die entwickelte systematische Lösung zielt darauf ab, eine nahtlose Pipeline bereitzustellen, die bei der Objekterkennung beginnt, durch Datenmodellzuordnung verbunden ist und bei der kontextspezifischen Bereitstellung von Anleitung endet.

4 Systemarchitektur und Lösungsansatz

Der folgende Abschnitt beschreibt die Module sowie die Systemarchitektur der entwickelten Lösung. Zunächst wird Modul I vorgestellt, das auf einem YOLO-basierten Objekterkennungsalgorithmus basiert und für die Erfassung der Bauteile in Echtzeit zuständig ist. Darauf aufbauend erläutert Modul II die Erstellung kontextspezifischer Montageanleitungen auf Basis einer semantischen Datenstruktur, die die Mitarbeiter gezielt durch den jeweiligen Arbeitsschritt führt. Abschließend wird der Informationsfluss des systematischen Lösungsansatzes beschrieben, der den Austausch und die Verknüpfung zwischen den Modulen sowie deren Integration in die Benutzeroberfläche veranschaulicht.

4.1 Modul I: YOLO-basierte Objekterkennung

Der YOLO-Algorithmus ist der Kern des Moduls I, das die Hauptfunktion der Objekterkennung übernimmt. Das Modul erkennt das Bauteil innerhalb der Montageumgebung und liefert die Bauteilinformationen als Eingabe für das Anleitungssystem. Um eine robuste Leistung zu erzielen, wird der YOLO-Algorithmus anhand synthetischer Bilddaten trainiert, die aus 3D-Modellen der entsprechenden Bauteile generiert werden. Dieser Ansatz ermöglicht es dem Modul, Bauteilmerkmale unter verschiedenen Perspektiven, Lichtverhältnissen und Ausrichtungen zu lernen, ohne dass umfangreiche manuelle Annotationen von Daten aus der realen Welt erforderlich sind. Durch die Nutzung virtueller Modelle kann der Trainingsdatensatz effizient skaliert und diversifiziert werden, wodurch eine Generalisierung auf die physische Montageumgebung gewährleistet ist.

Das YOLO-Modul ist mit dem in der Workstation installierten Kamerasystem verbunden. Die Kamera erfasst Echtzeitbilder der Umgebung des Mitarbeiters, die dann vom trainierten YOLO-Algorithmus verarbeitet werden. Der Algorithmus analysiert die Bilder in einem einzigen Durchlauf und gibt die Erkennungsergebnisse aus, bestehend aus Begrenzungsrahmen und Klassenbezeichnungen der identifizierten Bauteile. Diese Erkennungsergebnisse dienen als digitale Darstellung des physischen Montagezustands und ermöglichen es dem nachfolgenden Anleitungsmodul, den relevanten Montageschritt und die damit verbundenen Anleitungen zu bestimmen.

Durch diese Funktionalität stellt das Modul I für YOLO-basierte Objekterkennung eine direkte Verbindung zwischen der visuellen Realität der Montageumgebung und den strukturierten Datenmodellen her, die dem Anleitungssystem zugrunde liegen. Es stellt sicher, dass die Anleitungsinformation stets auf den tatsächlich am Arbeitsplatz vorhandenen Bauteilen basieren, und ermöglicht so eine kontextspezifische und adaptive Unterstützung für die Mitarbeiter, die Montagetätigkeiten durchführen.

4.2 Modul II: Montageanleitung

Das Montageanleitungsmodul bildet die zweite Kernkomponente des gezielten Systems. Die Funktion dieses Moduls besteht darin, dem Mitarbeiter kontextspezifische Anleitungen anzubieten, indem es erkannte Bauteile dynamisch mit strukturierten Anleitungsinformationen verknüpft. Das Modul basiert auf dem MMPD, in dem Datenklassen und ihre Zuordnung zu den Kategorien Objekt, Prozess, Ressource und Kontext vordefiniert sind. Diese strukturierte Grundlage gewährleistet Konsistenz, Erweiterbarkeit und semantische Klarheit bei der Darstellung von Montagewissen.

Innerhalb des ontologischen Bearbeitungswerkzeugs werden Instanzen nach den Definitionen dieser Datenklassen generiert. Beispielsweise kann eine Instanz in der Produktklasse den Namen und die Nummer des Bauteils angeben, während eine Instanz in der Werkzeugklasse die Verwendung eines Sechskantschlüssels definieren kann. In ähnlicher Weise beschreiben Maschinenklassen die erforderlichen Maschinen und Hilfsstoffe, und Mitarbeiterklassen erfassen mitarbeiterbezogene Attribute wie bevorzugte Sprache und Erfahrung. Durch die Abstimmung dieser Instanzen auf den entsprechenden Prozessschritt stellt das Modul sicher, dass jede Anleitung vollständig kontextualisiert und direkt auf die aktuelle Aufgabe anwendbar ist.

Das Modul empfängt die von Modul I mit der YOLO-basierten Objekterkennung generierten Erkennungsergebnisse, die die derzeit in der Montageumgebung vorhandenen Bauteile anzeigen. Diese Ergebnisse werden als Schlüssel für die Abfrage der ontologischen Datenstruktur verwendet, um den relevanten Prozessschritt und die zugehörigen Informationen abzurufen. Die Ausgabe wird dann zusammengestellt, das Produkt-, Ressourcen- und Kontextdetails in einem prägnanten Anleitungsformat integriert.

In der Praxis können die Anleitungsinformationen in verschiedenen Formaten formuliert werden. Der Ontologie-Editor Protégé kann zum Bearbeiten der Metadaten und zum Erstellen von Instanzen verwendet werden. Der Editor bietet auch verschiedene Formate zum Exportieren des instanziierten Datenmodells, zum Beispiel CSV-, XML-, und JSON-Datei – siehe ein Beispiel in **Bild 3**. Dies ermöglicht die Anwendbarkeit der Anleitungsinformationen in verschiedenen Implementierungsumgebungen und kann an die tatsächlichen Anforderungen angepasst werden.

Durch dieses Design bietet das Montageanleitungsmodul eine systematische und adaptive Methode zur Unterstützung der Mitarbeiter. Durch die Kombination von semantischer Wissensrepräs-

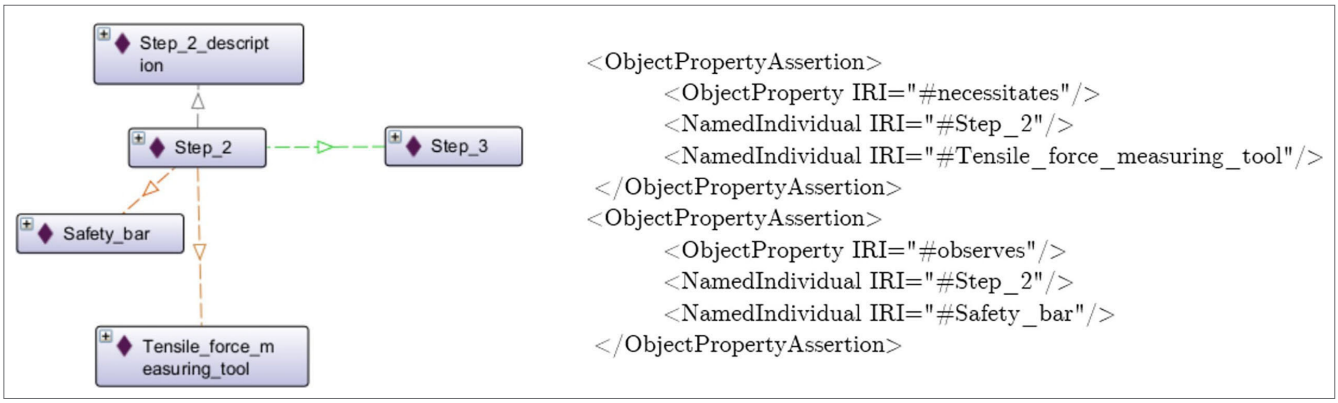

Bild 3 Darstellung von Metadaten in Protégé und der entsprechenden exportierten XML-Datei. *Grafik: RWTH Aachen*

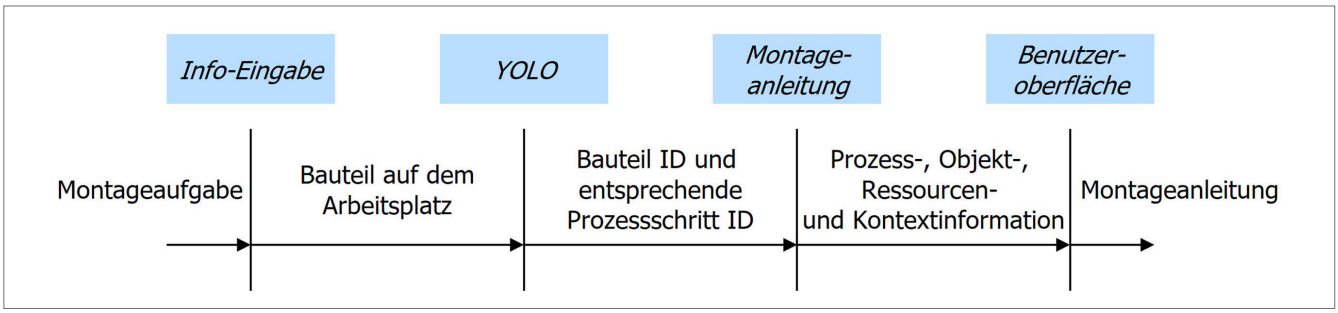

Bild 4 Informationsfluss zwischen verschiedenen Ebenen des Lösungsansatz. *Grafik: RWTH Aachen*

sentation mit Echtzeit-Objekterkennung wird sichergestellt, dass die dem Mitarbeiter präsentierten Informationen nicht nur umfassend, sondern auch situationsrelevant sind, wodurch die kognitive Arbeitsbelastung reduziert, und die Montagequalität verbessert wird.

4.3 Informationsfluss des systematischen Lösungsansatz

Das entwickelte System integriert das YOLO-Objekterkennungsmodul mit dem Montageanleitungsmodul. Es wird eine Pipeline von der Erfassung der realen Montageumgebung bis zur Bereitstellung kontextspezifischer Montageanleitung eingerichtet. In **Bild 4** wird der Informationsfluss dargestellt, der den Informationsaustausch zwischen den Ebenen zeigt. An diesem Lösungsansatz sind vier Ebenen beteiligt: Arbeitsplatz, YOLO-Objekterkennung, kontextspezifische Montageanleitungen und Benutzeroberfläche.

Der Informationsfluss beginnt mit der Informationseingabe vom Arbeitsplatz, wo das Kamerasystem Echtzeitbilder der Montageumgebung aufnimmt. Diese Bilder liefern eine visuelle Darstellung der aktuellen Umgebung, insbesondere der beteiligten Bauteile. Die Daten werden an das YOLO-Objekterkennungsmodul übertragen, das die Eingabe in einem einzigen Vorwärtsgang verarbeitet.

In der nächsten Phase analysiert das Objekterkennungsmodul die Bilder und gibt Erkennungsergebnisse aus, die aus Begrenzungsrahmen, Objektbewertungen und Klassenbezeichnungen für identifizierte Bauteile bestehen. Diesen identifizierten Bauteilen dienen als Anker für die Abfrage des Montageanleitungsmoduls.

Das Montageanleitungsmodul basiert auf dem MMPD, wobei vordefinierte Datenklassen die Kategorien Produkt, Prozess, Ressource und Kontext darstellen. Basierend auf den Erkennungsergebnissen des vorherigen Objekterkennungsmoduls ruft das Modul die entsprechenden Instanzen aus der Datenbank ab, in der die Anleitungsinformationen gespeichert sind. Beispielsweise wird ein erkannter Bauteil seiner Produktinstanz (Name, Nummer) zugeordnet, die wiederum mit dem zugehörigen Prozessschritt verknüpft ist. Dieser Prozessschritt ist weiter mit Ressourceninformationen (zum Beispiel erforderliches Werkzeug wie Sechskantschlüssel, erforderliche Maschine) und Kontextdaten (zum Beispiel Rolle des Mitarbeiters, Umgebungsparameter) verbunden.

Schließlich kompiliert das System in der Ausgabe die abgerufenen Informationen zu einer kontextspezifischen Anleitung, die auf der Benutzeroberfläche dargestellt wird. Die Benutzeroberfläche enthält nur die für den aktuellen Schritt relevanten Details – wie die identifizierten Bauteile, das erforderliche Werkzeug und den nächsten Montagevorgang – wodurch der kognitive Aufwand für den Mitarbeiter minimiert wird. Die Anleitungen können über beliebige Schnittstellen (zum Beispiel Monitor, Tablet oder AR-Gerät) angezeigt werden, wodurch sichergestellt wird, dass die Anleitungen sowohl anpassungsfähig als auch situationsbezogen sind.

Durch die Verknüpfung von Echtzeiterkennung mit einer strukturierten Anleitungsinformation schafft das System einen geschlossenen Informationsfluss: visuelle Eingabe → Objekterkennung → Abruf von Anleitungen → Mitarbeiterunterstützung. Dieser Arbeitsablauf reduziert direkt den Bedarf des Mitarbeiters, allgemeine Handbücher zu interpretieren oder variantenbezogene Dokumentationen zu durchsuchen.

5 Fazit und Ausblick

Diese Arbeit stellt die Entwicklung einer Pipeline zur Erstellung der Montageanleitungen vor, die Computer Vision-gestützte

Objekterkennung mit MMPD-basierten Montageanleitungsmodellierung integriert. Die Grundidee besteht darin, die Lücke zwischen der physischen Montageumgebung und dem virtuellen Informationsraum zu schließen, indem der YOLO-Algorithmus für die Echtzeit-Erkennung von Bauteilen eingesetzt und die Ergebnisse auf strukturierte Anleitungsinformationen abgebildet werden, die aus dem MMPD abgeleitet werden. Durch diese Integration bietet das System kontextspezifische Anleitungen, die auf die erkannten Objekte, die zugehörigen Prozessschritte und die erforderlichen Ressourcen zugeschnitten sind. Durch die systematische Kombination von Wahrnehmung und semantischer Modellierung begegnet das Framework den Herausforderungen der kognitiven Arbeitsbelastung bei der manuellen Montage und schafft die Grundlage für ein adaptives, mitarbeiterzentriertes Anleitungssystem.

Die konzeptionelle Pipeline veranschaulicht zwar erfolgreich die Integration von Erkennung und Anleitung, doch sind weitere Entwicklungen erforderlich, um den Ansatz zu validieren und zu verfeinern. Zunächst muss die Pipeline an einer kompletten Produktmontage getestet werden, um die Leistung über einen End-to-End-Prozess hinweg und nicht nur an isolierten Komponenten zu bewerten. Anschließend soll die Robustheit und Durchführbarkeit der Pipeline anhand komplexerer Produkte bewertet werden, bei denen eine höhere Variabilität der Teile, Montageabläufe und Kontextfaktoren die Anpassungsfähigkeit des Systems herausfordern wird. Eine solche Validierung wird wichtige Erkenntnisse über die Skalierbarkeit, Verallgemeinerbarkeit und Integration in reale industrielle Umgebungen liefern.

FÖRDERHINWEIS

Diese Arbeit ist Teil des Forschungsprojekts „DiCES", das vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) im Rahmen der GreenTech-Förderlinie „Entwicklung digitaler Technologien" unter dem Förderkennzeichen 01MN23022E gefördert und vom Deutschen Zentrum für Luft- und Raumfahrt (DLR) unterstützt wird. Für den Inhalt sind die Autoren verantwortlich.

LITERATUR

[1] Breznik, Matic ; Buchmeister, Borut ; Vujica Herzog, Nataša: Evaluation of the EAWS Ergonomic Analysis on the Assembly Line: Xsens vs. Manual Expert Method— A Case Study. In: Sensors Bd. 25, MDPI AG (2025), Nr. 15, S. 4564

[2] Brolin, A.; Thorvald, P. ; Case, K.: Experimental study of cognitive aspects affecting human performance in manual assembly. In: Production & Manufacturing Research Bd. 5, Informa UK Limited (2017), Nr. 1, S. 141–163

[3] Biondi, F. N. ; Cacanindin, A. ; Douglas, C. ; Cort, J.: Overloaded and at Work: Investigating the Effect of Cognitive Workload on Assembly Task Performance. In: Human Factors: The Journal of the Human Factors and Ergonomics Society Bd. 63, SAGE Publications (2020), Nr. 5, S. 813–820

[4] Berlin, C.; Bergman Wollter, M.; Chafi, Maral Babapour ; Falck, A.-C. ; Örtengren, R.: A Systemic Overview of Factors Affecting the Cognitive Performance of Industrial Manual Assembly Workers. In: Lecture Notes in Networks and Systems: Springer International Publishing, 2021. ISBN 9783030746070, S. 371–381

[5] Zhou, L.; Zhang, L.; Konz, N.: Computer Vision Techniques in Manufacturing. In: IEEE Transactions on Systems, Man, and Cybernetics: Systems Bd. 53, Institute of Electrical and Electronics Engineers (IEEE) (2023), Nr. 1, S. 105–117

[6] Qamar, R.; Zardari, B. A.: Application of Computer Vision in Manufacturing. In: Machine Vision and Industrial Robotics in Manufacturing: CRC Press, 2024, S. 36–56

[7] Zhou, Longfei ; Zhang, Lin ; Konz, Nicholas: Computer Vision Techniques in Manufacturing. In: IEEE Transactions on Systems, Man, and Cybernetics: Systems Bd. 53, Institute of Electrical and Electronics Engineers (IEEE) (2023), Nr. 1, S. 105–117

[8] Ettalibi, Abdelfatah ; Elouadi, Abdelmajid ; Mansour, Abdeljebar: AI and Computer Vision-based Real-time Quality Control: A Review of Industrial Applications. In: Procedia Computer Science Bd. 231, Elsevier BV (2024), S. 212–220

[9] Raisul Islam, Md ; Zakir Hossain Zamil, Md ; Eshmam Rayed, Md ; Mohsin Kabir, Md ; Mridha, M. F. ; Nishimura, Satoshi ; Shin, Jungpil: Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. In: IEEE Access Bd. 12, Institute of Electrical and Electronics Engineers (IEEE) (2024), S. 121449–121479

[10] Pathak, Ajeet Ram ; Pandey, Manjusha ; Rautaray, Siddharth: Application of Deep Learning for Object Detection. In: Procedia Computer Science Bd. 132, Elsevier BV (2018), S. 1706–1717

[11] Jiang, P.; Ergu, D. ; Liu, F. ; Cai, Y.; Ma, B.: A Review of Yolo Algorithm Developments. In: Procedia Computer Science Bd. 199, Elsevier BV (2022), S. 1066–1073

[12] Redmon, J. ; Divvala, S. ; Girshick, R.; Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) : IEEE, 2016, S. 779–788

[13] Aboyomi, Dalmar Dakari ; Daniel, Cleo: A Comparative Analysis of Modern Object Detection Algorithms: YOLO vs. SSD vs. Faster R-CNN. In: ITEJ (Information Technology Engineering Journals) Bd. 8, IAIN Syekh Nurjati Cirebon (2023), Nr. 2, S. 96–106

[14] Kang, S.; Hu, Z.; Liu, L.; Zhang, K.; Cao, Z.: Object Detection YOLO Algorithms and Their Industrial Applications: Overview and Comparative Analysis. In: Electronics Bd. 14, MDPI AG (2025), Nr. 6, S. 1104

[15] Claeys, A.; Hoedt, S.; Landeghem, H. Van; Cottyn, J.: Generic Model for Managing Context-Aware Assembly Instructions. In: IFAC-PapersOnLine Bd. 49, Elsevier BV (2016), Nr. 12, S. 1181–1186

[16] Josifovska, K.; Yigitbas, E.; Engels, G.: A Digital Twin-Based Multi-modal UI Adaptation Framework for Assistance Systems in Industry 4.0. In: International Conference on Human-Computer Interaction, Springer, Cham (2019), S. 398-409

[17] Bao, J.; Guo, D.; Li, J.-, Zhang, J.: The modelling and operations for the digital twin in the context of manufacturing. In: Enterprise Information Systems Bd. 13, Informa UK Limited (2018), Nr. 4, S. 534–556

[18] Claeys, A. ; Hoedt, S. ; Schamp, M. ; Landeghem, H. Van ; Cottyn, J.: Ontological Model for Managing Context-aware Assembly Instructions. In: IFAC-PapersOnLine Bd. 51, Elsevier BV (2018), Nr. 11, S. 176–181

[19] Nuy, L.; Rotering, J.; Rachner, J.; Kiesel, R.; Schmitt, R. H.: Conception of a data model for a digital twin for context-specific work instructions. In: Procedia CIRP Bd. 118, Elsevier BV (2023), S. 312–317

[20] Cramer, S.; Hoffmann, M.; Schlegel, P.; Kemmerling, M.; Schmitt, R. H.: Towards a flexible process-independent meta-model for production data. In: Procedia CIRP Bd. 99, Elsevier BV (2021), S. 586–591

[21] Liang, J.; Pelzer, L.; Müller, K. ; Cramer, S.; Greb, C.; Hopmann, C. ; Schmitt, R. H.: Towards predictive quality in production by applying a flexible process-independent meta-model. In: Procedia CIRP Bd. 104, Elsevier BV (2021), S. 1251–1256

Junjie Liang

Alexander Moriz

Dr. Amon Göppert

Prof. Robert H. Schmitt

Werkzeugmaschinenlabor WZL der RWTH Aachen
Lehrstuhl für Informations-, Qualitäts- und Sensorsysteme in der Produktion
Campus Boulevard 30, 52074 Aachen
www.wzl.rwth-aachen.de