Sachbericht

Zuwendungsempfänger:	Förderkennzeichen:
TU Dortmund	**28DK102F20**

Vorhabenbezeichnung:

Verbundprojekt: Ressourceneffiziente KI für eingebettete Systeme in Landmaschinen (resKIL) - Teilprojekt F

Laufzeit des Vorhabens:

01.01.2021 - 30.06.2024

Berichtszeitraum:

01.01.2021 - 30.06.2024

INHALT

TEIL I: Kurzbericht 2

Aufgabenstellung + wiss./technischer Stand 2

Planung und Ablauf des Vorhabens 3

Wesentliche Ergebnisse 3

Zusammenarbeit mit anderen Stellen 3

Teil II: Eingehende Darstellung 4

Durchgeführte Arbeiten und Ergebnisse 4

Wichtigsten Positionen des zahlenmäßigen Nachweises 7

Geleistete Arbeit sowie deren Notwendigkeit und Angemessenheit 7

Voraussichtliche Nutzen 7

Bekannt gewordener Fortschritt bei anderen Stellen 8

Erfolgte und geplante Veröffentlichungen 8

TEIL I: KURZBERICHT

AUFGABENSTELLUNG + WISS./TECHNISCHER STAND

Das Projekt resKIL hatte das Ziel, durch eine detaillierte Analyse und Generalisierung von Problemstellungen sowie durch die Entwicklung einer anpassungsfähigen Softwarearchitektur und einer KI-Toolchain, die Ressourcennutzung sowohl im personellen als auch im technischen Bereich zu optimieren. Dies sollte zu einer skalierbaren und variablen Lösung führen, die eine breitere Anwendung von maschinellem Lernen in der Landwirtschaft ermöglicht. Die Entwicklungen wurden am Beispiel der Qualitätsbestimmung von Erntegut und der Merkmalserkennung im Maschinenumfeld in Funktionsmuster integriert und umgesetzt. Fokussiert wurde dabei der effiziente Softwareentwurf von KI-Systemen, effiziente KI-Hardware, effiziente Datengewinnung, effiziente Datenaufbereitung und -labeling und effiziente Berechnung von KI-Algorithmen. Der Projektverlauf orientierte sich an den drei Ernteperioden, die den zeitlichen Ablauf und die Iterationen der Arbeitspakete bestimmten. Ein wesentlicher Bestandteil war zudem der Transfer der Entwicklungen über mehrere Ernten hinweg, um die Generalisierbarkeit und darüber hinaus eine Steigerung der Robustheit und Qualität der Ansätze nachweisen zu können.

Die TU Dortmund baute hierbei auf folgenden wissenschaftlichen und technischen Stand auf: Da die Daten nur zu bestimmten Zeiten im Jahr erhoben werden können, ist eine effiziente Stichproben- und Versuchsplanung entscheidend, um eine ausreichende Datenvariation sicherzustellen. Neben der Abdeckung weitreichender Testszenarien spielte die daraus resultierende Datenqualität und deren Beurteilung eine zentrale Rolle. Dabei bezieht sich die Datenqualität nicht nur auf die Messgenauigkeit, sondern auch auf die Relevanz und Vollständigkeit der Daten. Eine umfassende Testumgebung, die diese Aspekte integriert, war zu Projektbeginn in der wissenschaftlichen und praktischen Literatur nicht zufriedenstellend beschrieben. Im Rahmen des Projekts wurde die Kombination bestehender Ansätze zur Versuchsplanung (wie fraktionierte faktorielle Designs und sequentielle Versuchsplanung) mit der Anwendung von Datenqualitätsmetriken in einem iterativen Datengenerierungsprozess verfolgt. Imputationsmethoden sollten problemabhängig gewählt werden, um ungenaue Vorhersagen zu vermeiden. Die gemeinsame Vorhersage mehrerer Qualitätsmerkmale wurde empfohlen, um unplausible Ergebnisse zu verhindern. Zur Evaluation der Algorithmen sollten passende Metriken gewählt werden, die sowohl quantitative als auch qualitative Kriterien berücksichtigen und an die spezifischen Probleme angepasst sind.

Planung und Ablauf des Vorhabens

Es wurden 7 Arbeitspakete definiert, die aufeinander aufbauten und gemeinsam im Konsortium vorangetrieben wurden. Die TU Dortmund war aktiv in den Arbeitspaketen AP1 (Statistisches Framework zur Datenakquise und -aufbereitung), AP4 (Skalierbare KI-Plattform), AP6 (Evaluation) und AP7 (Projektkoordination) beteiligt.

Darüber hinaus wurden im Konsortium 3 gemeinsame Meilensteine definiert

- M1 (nach 12 Monaten): Auf Basis der statistischen Versuchsplanung wurden in der ersten Erntekampagne umfassende Trainingsdaten gewonnen. Umsetzungskonzepte für Framework, Architektur und Plattform wurden erstellt.
- M2 (nach 24 Monaten): Die aufgenommenen Daten wurden mit dem entwickelten Tool annotiert und der quantitative Nachweis für eine Reduktion des Annotationsaufwands wurde erbracht. Framework, Architektur und Plattform wurden als Funktionsmuster/Demonstrator vorgelegt und in der zweiten Ernte erprobt.
- M3 (nach 36 Monaten): Der Funktionsnachweis für die Anwendung von KI-Verfahren zur Ausführung auf eingebetteten Systemen wurde erbracht. Die Skalierbarkeit und Generalisierbarkeit der Methoden wurden durch verschiedene Anwendungen sowohl auf der Landmaschine als auch in der Cloud gezeigt.

Wesentliche Ergebnisse

Die Teilarbeiten der TU Dortmund konnten erfolgreich und weitgehend planungskonform durchgeführt werden. Alle Arbeiten zu den einzelnen Meilensteinen wurden fristgerecht erfüllt, wodurch der Funktionsnachweis für die Anwendung von KI-Verfahren auf eingebetteten Systemen erbracht werden konnte.

Zusammenarbeit mit anderen Stellen

Neben TU Dortmund setzt sich das Projekt resKIL aus den Partnern Claas CSE und CES, Zauberzeug, Uni Osnabrück und DFKI zusammen. Eine Zusammenarbeit ist je nach Fokus der APs mit den jeweiligen Partnern durchgeführt worden. Es wurden im Konsortialverbund monatliche Treffen durchgeführt, in denen die neuesten Ergebnisse ausgetauscht wurden.

TEIL II: EINGEHENDE DARSTELLUNG

DURCHGEFÜHRTE ARBEITEN UND ERGEBNISSE

Die TU Dortmund hat in AP1 (Statistisches Framework zur Datenakquise und -aufbereitung), AP4 (Skalierbare KI-Plattform), AP6 (Evaluation) und AP7 (Projektkoordination) aktiv mitgewirkt. Diese Aufgaben wurden über die gesamte Projektlaufzeit hinweg geplant und erfolgreich umgesetzt. Im Konsortium wurden zudem drei gemeinsame Meilensteine definiert, um die Fortschritte zu überwachen:

- M1 (nach 12 Monaten): Auf Basis der statistischen Versuchsplanung wurden in der ersten Erntekampagne umfassende Trainingsdaten gewonnen. Außerdem lagen Umsetzungskonzepte für Framework, Architektur und Plattform vor.

- M2 (nach 24 Monaten): Die aufgenommenen Daten wurden mit dem entwickelten Tool annotiert und der quantitative Nachweis für eine Reduktion des Annotationsaufwands erbracht. Framework, Architektur und Plattform lagen als Funktionsmuster/Demonstrator vor und wurden in der zweiten Ernte erprobt.

- M3 (nach 36 Monaten): Der Funktionsnachweis für die Anwendung von KI-Verfahren zur Ausführung auf eingebetteten Systemen wurde erbracht. Die Skalierbarkeit und Generalisierbarkeit der Methoden wurden durch verschiedene Anwendungen sowohl auf der Landmaschine als auch in der Cloud gezeigt.

Im ersten Projektjahr lag der Fokus der TU Dortmund auf der Entwicklung erster Stichproben- und Versuchspläne zur umfassenden Datenauswahl, -aufbereitung und -auswertung, mit dem Ziel, den KI-Trainingslösungsraum bestmöglich abzudecken. Dies wurde in enger Zusammenarbeit mit CLAAS unter Nutzung der spezifischen Erfahrungen und Expertise der Partner erreicht. In mehreren Workshops konnten die Ziele der einzelnen Anwendungsfälle sowie die Bedeutung relevanter Einflussfaktoren und deren Wechselwirkungen präzise definiert werden. Die Erstellung der Versuchspläne erfolgte in einem zweistufigen Verfahren. Die auf der Basis der Erfahrungen von CLAAS erstellten Pläne wurden statistisch analysiert und erweitert, um eine möglichst gute Abdeckung des Lösungsraums zu erreichen. Zusätzlich fanden Versuche auf Testparzellen statt, um Unsicherheiten während der Ernte, wie Probleme bei der Datenerfassung oder suboptimale Versuchsbedingungen, zu minimieren.

Ein weiteres Ziel war die Bewertung der Datenqualität der Erntedaten aus dem Jahr 2020 im Anwendungsfall Häcksler. Hierzu wurden die Daten zunächst statistisch analysiert, um genauere Einblicke in die Zusammenhänge zu gewinnen. Zudem wurde die Leistungsfähigkeit

des entwickelten KI-Algorithmus zur optischen Berechnung des Corn Silage Processing Score (CSPS) im Vergleich zur konventionellen, zeit- und kostenintensiven Laborbestimmung evaluiert. Diese Analyse umfasste auch die Bewertung der Stabilität und Performance des Algorithmus.

Im Laufe des ersten Jahres wurden zahlreiche gemeinsame Workshops durchgeführt, die nicht nur die oben genannten Fragestellungen behandelten, sondern auch die Entwicklung von Konzepten für die KI-Plattform und die Diskussion über die geeigneten Evaluationsmetriken in den einzelnen Anwendungsfällen, um eine präzise und objektive Bewertung der entwickelten Methoden zu ermöglichen. Diese Workshops dienten zudem dazu, die besonderen Herausforderungen und Randbedingungen der Anwendungsdomäne eingehend zu analysieren.

Durch die beschriebenen Arbeiten konnte die TU Dortmund nach 12 Projektmonaten alle Arbeiten, die im Zusammenhang mit dem Erreichen von Meilenstein M1 standen, vollständig und fristgerecht abschließen.

Im zweiten Jahr des Projekts wurden auf den Ergebnissen des ersten Jahres aufgebaut und eine weitergehende Analyse der Stichproben- und Versuchspläne vorgenommen. Ziel war es, die bereits entwickelten Ansätze zur Datenauswahl und -auswertung weiter zu verfeinern, wobei erneut eng mit CLAAS zusammengearbeitet wurde. In einer Reihe gemeinsamer Workshops wurden die Ziele der einzelnen Use Cases erneut diskutiert und die Bedeutung bestimmter Einflussfaktoren sowie deren Wechselwirkungen und Skalenbereiche angepasst. Die Versuchspläne des Vorjahres wurden umfassend analysiert und auf Basis der Erfahrungen von der ersten Ernte und den Ergebnissen der Analyse für die nächste Ernte in einem zweistufigen Verfahren überarbeitet. Auch in diesem Jahr fanden erneut Feld- und Erntefahrten sowie Versuche auf Testparzellen statt, um Unsicherheiten während der Ernte bestmöglich zu minimieren.

Ein weiteres Ziel bestand darin, die Qualität der Erntedaten aus dem ersten Jahr im Anwendungsfall Häcksler zu bewerten. Dazu wurden die Daten zunächst wieder statistisch analysiert, um genauere Einblicke in die Zusammenhänge zu erhalten. Auf Grundlage dieser Analyse setzte die TU Dortmund die Erstellung und Weiterentwicklung von statistischen Modellen fort, um zu untersuchen, wie gut die optische CSPS-Bestimmung im Vergleich zur konventionellen Laborbestimmung abschneidet. Zudem wurde der Vergleich der Performance des aktuell verwendeten KI-Algorithmus zur optischen CSPS-Berechnung mit alternativen statistischen Methoden der Regressionsanalyse begonnen, die alle aus der Bildsegmentierung gewonnen Informationen als Eingabe verwenden. In dieser Analyse haben wir den Trade-off zwischen der genauen CSPS-Bestimmung und der damit verbundenen Ressourcenallokation

untersucht, um ein besseres Verständnis für die Praktikabilität und Effektivität dieser ressourceneffizienten Ansätze zu gewinnen.

Die Arbeiten der TU Dortmund in Bezug auf Meilenstein M2 nach Projektmonat 24 wurden vollständig erfüllt. Im dritten Jahr des Projektes wurden die bisher gewonnenen Erkenntnisse zielgerichtet weiterverfolgt.

Anhand der Ergebnisse aus den ersten beiden Jahren konzentrierte sich die TU Dortmund darauf, die Stichproben- und Versuchspläne sequenziell weiterzuentwickeln, zu analysieren und zu verbessern, um eine umfassende Datenauswahl, -aufbereitung und -auswertung zu ermöglichen und somit den KI-Trainingslösungsraum optimal abzudecken. Die enge Zusammenarbeit mit CLAAS wurde fortgeführt, und die spezifischen Expertisen beider Partner sowie die gesammelten Erfahrungen der ersten beiden Projektjahre flossen weiterhin in den Prozess ein. Die Versuchspläne für die Ernten im Jahr 2023 wurden wieder in einem zweistufigen Verfahren erstellt und erweitert, um den Lösungsraum möglichst umfassend abzudecken.

Ein weiteres Ziel bestand darin, im Anwendungsbereich Häcksler die Qualität der Erntedaten aus dem zweiten Jahr zu evaluieren und zu bewerten. Dabei wurde eine vergleichbare statistische Analyse (u.a. mittels verschiedener statistischer Lernverfahren) wie im Vorjahr durchgeführt, um unter anderem eine Kontrolle über die Konsistenz der Ergebnisse zu gewährleisten. Basierend auf diesen Analysen wurde die Weiterentwicklung statistischer Methoden fortgesetzt, um die Effizienz der optischen CSPS-Bestimmung im Vergleich zum traditionellen Laborverfahren zu evaluieren. Parallel dazu wurden verschiedene statistische Algorithmen zur Bestimmung des CSPS-Wertes anhand von segmentierten Erntebildern weiter untersucht.

Darüber hinaus unterstützte die TU Dortmund die anderen Projektpartner bei statistischen Fragestellungen zur Evaluierung. Dies beinhaltete unter anderem die statistische Unterstützung des DFKI bei der Planung von Versuchen zur Bewertung verschiedener Annotationsmethoden.

In den sechs Monaten der kostenneutralen Verlängerung wurde neben Nachanalysen zudem an zwei wissenschaftlichen Veröffentlichungen gearbeitet, die durch die Arbeiten am resKIL Projekt motiviert sind. Die Arbeiten befassen sich mit der ressourceneffizienten Berechnung von baumbasierten Lernverfahren sowie der ML-basierten Zeitreihenprognose und befinden sich aktuell in den finalen Schritten.

Wichtigste Positionen des zahlenmäßigen Nachweises

Der Verlauf der Arbeiten im Projekt der TU Dortmund folgte weitestgehend der im Projektantrag formulierten Planung. Alle im Arbeitsplan für die TU Dortmund formulierten Aufgaben wurden erfolgreich nach einer sechsmonatigen kostenneutralen Verlängerung bearbeitet, es waren keine zusätzlichen Ressourcen für das Projekt nötig.

Die starke Fokussierung auf den Demonstrator, die im initialen Plan unterschätzt wurde, sowie das Fehlen eines Projektmitarbeiters zu Beginn des Projekts bei unserem Partner führten zu Verzögerungen. Diese Umstände, zusammen mit der hohen Interdisziplinarität, machten die beantragte Verlängerung notwendig. Bei der TU Dortmund konzentrierte sich diese Verlängerung insbesondere auf die statistische Nachanalyse und Verwertung der wissenschaftlichen Erkenntnisse aus der gesamten Projektlaufzeit.

Geleistete Arbeit sowie deren Notwendigkeit und Angemessenheit

Die definierten Projektziele wurden im Rahmen des Projekts vollständig erreicht. Die geförderten Maßnahmen waren sowohl notwendig als auch in der Höhe angemessen, da sie eine umfassende Analyse der Problemstellung, die Entwicklung eines integrierten Gesamtsystems und die Integration in einen realen Demonstrator ermöglichten. Darüber hinaus war die Einbindung der TU Dortmund in das Forschungsprojekt mit einem zusätzlichen Personalbedarf verbunden, der nicht aus eigenen Mitteln gedeckt werden konnte.

Voraussichtliche Nutzen

Die Arbeiten der TU Dortmund konzentrierten sich über die verschiedenen Arbeitspakete hinweg auf die statistische Expertise, beginnend bei der Datenerhebung bis hin zur finalen Evaluation. Das entwickelte Vorgehen sowie die eingesetzten Methoden können als solide Grundlage für zukünftige Entwicklungen dienen. Zudem lassen sich die im Projekt gewonnenen Erkenntnisse zur effizienten Entwicklung von KI-basierten Lösungen direkt in zukünftige Projekte übertragen.

Zusätzlich fließen die Ergebnisse des Projekts in die Lehre und Ausbildung von Studierenden ein. Die entwickelten Methoden und Tools bieten praxisnahe Anwendungsbeispiele, die in Lehrveranstaltungen genutzt werden können, um Studierende auf die Herausforderungen der Datenanalyse und KI-Entwicklung vorzubereiten.

Bekannt gewordener Fortschritt bei anderen Stellen

Es sind während der Projektlaufzeit keine Ergebnisse von dritter Seite bekannt geworden, die für die Durchführung des Vorhabens relevant waren. Entsprechend waren keine Änderungen der Zielsetzung des Vorhabens notwendig.

Erfolgte und geplante Veröffentlichungen

Es sind zwei Veröffentlichungen geplant. Die Papiere wurden im September bzw. werden voraussichtlich im November 2024 eingereicht.