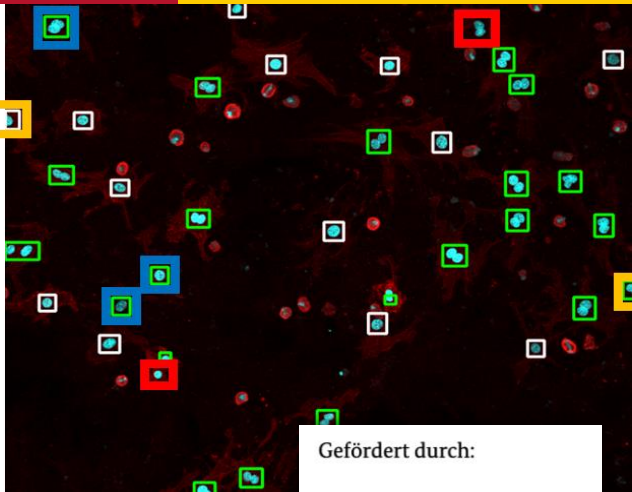
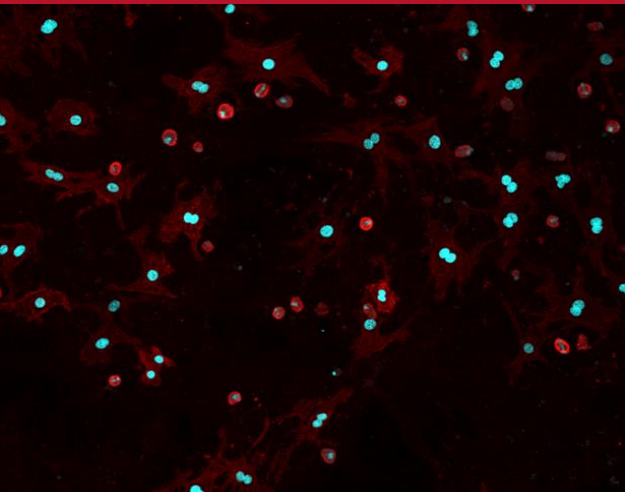


Industrielle Gemeinschaftsforschung

F.O.M.

025



Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Projektinformationen

IGF-Nr.:	21361 N
Laufzeit:	09/2020 - 08/2022
Fördersumme:	248.918 EUR
Industrieleistungen:	73.800 EUR

Forschungseinrichtungen

- Fraunhofer-Institut für Produktionstechnologie IPT Aachen

Projektbegleitender Ausschuss

- ALS Automated Lab Solutions GmbH ^{KMU}
- Bayer AG
- Cellmatiq GmbH ^{KMU}
- CytoSmart
- IconPro GmbH ^{KMU}
- Labforward GmbH ^{KMU}
- MABRI.VISION GmbH ^{KMU}
- MINDPEAK GmbH ^{KMU}
- Olympus Soft Imaging Solutions GmbH
- ORACLE Deutschland B.V. & Co. KG
- PicoQuant GmbH ^{KMU}
- Ruhruniversität Bochum
- Stammzellnetzwerk NRW e. V.
- Taorad GmbH ^{KMU}
- Uniklinik Köln
- Uniklinik RWTH Aachen

- **Automatisierte Auswertung bio-mediz. Bilddaten von Zellkulturen**
- **Automat. Auswahl u. Konfigurierung von Deep Learning-Modellen**

Cell Culture Analysis Tool (AixCell)

Die Herausforderung

Die zeitaufwendige Analyse biomedizinischer Bilddaten, z. B. in der Zellmikroskopie, wird bisher von spezialisiertem Labpersonal durchgeführt und ist dementsprechend kostenintensiv. Die Subjektivität der Auswertung und die Anfälligkeit für anwendungs- und gerätespezifische Fehler beeinträchtigen zudem die Vergleichbarkeit der Ergebnisse.

Methoden des Deep Learning (DL), die tiefe künstliche neuronale Netze zur semantischen Wissensextraktion aus Bilddaten nutzen, ermöglichen eine Hardware-unabhängige automatisierte und objektive Bildauswertung. Die Entwicklung solcher Anwendungen erfordert jedoch Expertise und Erfahrung in den Fachgebieten Datenwissenschaften, maschinelles Lernen, Informationstechnologie und Software-Entwicklung. Über das IT-Knowhow im benötigten Umfang und in ausreichender Tiefe verfügen z. B. Analysen-Spezialisten oder Mediziner nur selten, um funktionsfähige DL-Anwendungen für ihre jeweiligen konkreten Problemstellungen selbst entwickeln zu können. Die Entwicklung spezifischer Einzellösungen bei externen IT-Büros in Auftrag zu geben, ist aus wirtschaftlichen Gründen oft nicht umsetzbar.

Die Innovationsidee

Ziel des Projekts war, eine intuitiv nutzbare Software zur ganzheitlichen DL-basierenden Bildauswertung zu entwickeln, inklusive automatisierter Vorverarbeitung der Daten, Algorithmus-Auswahl und Konfiguration von DL-Modellen, die biomedizinische Experten für verschiedene Zell- und Gewebeanalysen auch ohne umfangreiche IT-Kenntnisse eigenständig nutzen können. Hierzu sollte mithilfe einer domänenspezifischen selbstlernenden Entscheidungslogik (engl. Automated Machine Learning; AutoML) die Konfiguration einer aus Datenvorverarbeitung, neuronaler Netzarchitektur, Lernalgorithmus und Nachbearbeitung bestehenden Datenpipeline für die zur Verfügung stehenden Rechenressourcen und den spezifischen Anwendungsfall optimiert werden. Die Datenpipeline soll im Anschluss trainiert und dem biomedizinischen Experten zur Nutzung und Integration in dessen Analyse-Workflow zur Verfügung gestellt werden.

Die Ergebnisse

Um den komplexen und iterativen Entwicklungsprozess einer Datenpipeline zu automatisieren, wurde ein domänenspezifisches AutoML-System basierend auf

Projektbegleitende akademische Abschlussarbeiten

- Drei Bachelor-Arbeiten
- Sieben Master-Arbeiten
- Eine Promotion

Das Programm "Industrielle Gemeinschaftsforschung" (IGF) ...

... fördert Studien zur industriellen Machbarkeit von Innovationsideen und beschleunigt so Technologietrends. Dazu arbeiten Wissenschaft, Industrie und Politik zusammen:

0 Das **BMWK** fördert vorwettbewerbliche, innovationsorientierte Forschung mit dem IGF-Programm.

1 **Industrie und Wissenschaftler** entwickeln Innovationsideen und geben Projekimpulse.

2 **IGF-Forschungsvereinigungen**, wie die F.O.M., finden Forschungspartner.

3 **Wissenschaftler** von je 1-3 Forschungseinrichtungen schreiben Förderanträge.

4 **Industrieunternehmen** beraten bei der Entwicklung der Anträge.

5 Die **Forschungsvereinigungen** optimieren die Qualität der Vorhaben und der Anträge und reichen die Anträge ein.

6 Die **Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF)** lässt die Anträge durch **Experten aus Industrie und Wissenschaft** begutachten.

7 Das **BMWK** finanziert die Forschungskosten bis max. 275/525/750 T EUR.

8 Die **Industrie** teilt sich die Administrationskosten.

9 Die **Wissenschaftler** der Forschungseinrichtungen führen die Forschung durch.

10 Die **Forschungsvereinigungen** stellen mithilfe von je 10-20 Unternehmen projektbegleitender **Industrieausschüsse** mit mindestens 50 % KMU einen regen Technologietransfer bis in die Branchen hinein sicher.

11 Die **Industrie** sorgt durch Bekundung ihrer Interessen für die Praxisrelevanz der Forschung, steuert Industrieexpertise bei und validiert die Ergebnisse.

Gemeinsam stärken wir die Innovationskraft des Mittelstands und den Fachkräftenachwuchs in Deutschland.

Für eine ausführlichere Fassung des Abschlussberichts wenden Sie sich bitte an:

Kontakt / Impressum

Forschungsvereinigung F.O.M.
Werderscher Markt 15, 10117 Berlin
030 4140 21-50

info@forschung-fom.de
www.forschung-fom.de



semantischer Segmentierung von Mikroskopiedaten entwickelt. Die dazu benötigte modulare Softwarebibliothek, Daten- und Wissensbasis wurden anhand von zehn verschiedenen Anwendungsfällen aufgebaut.

Das finale AutoML-System besteht aus fünf zentralen Komponenten: 1) Ein vordefinierter Suchraum, 2) eine Wissensbasis aus Metadaten zu den Merkmalen der Anwendungsfälle und der Modell-Performance, 3) eine Datenvorverarbeitung zur Erschließung des maximal extrahierbaren Informationsgehalts, 4) ein Meta-Learning-Modell zur Bestimmung einer Rangfolge der Datenpipelines im Hinblick auf ihre Eignung für die Bearbeitung der Analyseaufgabe und 5) ein Multi-Fidelity-Ansatz zur Identifizierung der am besten performenden Datenpipeline.

Das AutoML-System wurde in eine lokal nutzbare und in eine Cloud-basierte Anwendung integriert. Bei letzterer kann über einen Internetbrowser eine neue Analyseaufgabe definiert, Daten hochgeladen, diese annotiert und das AutoML-System gestartet werden. Ausgegeben werden die Konfiguration der optimalen Datenpipeline und eine Bewertung der Auswahl. Das Modell kann nun in der Laborumgebung direkt zur Zellkulturanalyse genutzt oder in einen automatisierten Analyse-Workflow integriert werden.

Die Performance der durch das AutoML-System für verschiedene spezielle Anwendungsfälle konfigurierten Datenpipelines liegt nahe bei individuell entwickelten Modellen und bestätigt damit die Funktionalität des Ansatzes einer vollständigen Automatisierung der DL-Entwicklung und Bildauswertung in abgegrenzten Domänen, ohne gravierende Nachteile in den Modellgenauigkeiten.

Aktuell erfordern einige Analysen spezifische Nachbearbeitungsschritte, mithilfe derer die Segmentationsmasken der neu-

ronalen Netze auf z. B. morphologische Eigenschaften hin untersucht werden können. Die modulare Software-Struktur erlaubt jedoch, erforderliche Nachbearbeitungen mithilfe einfacher Ergänzungen zu implementieren. Mit steigender Anzahl solcher Implementierungen wird die Anzahl zur Verfügung stehender standardisierter Ergänzungsmodulare und dadurch auch der Funktionsumfang der Anwendung zunehmen.

Die Verwertung

KMU-Nutzen

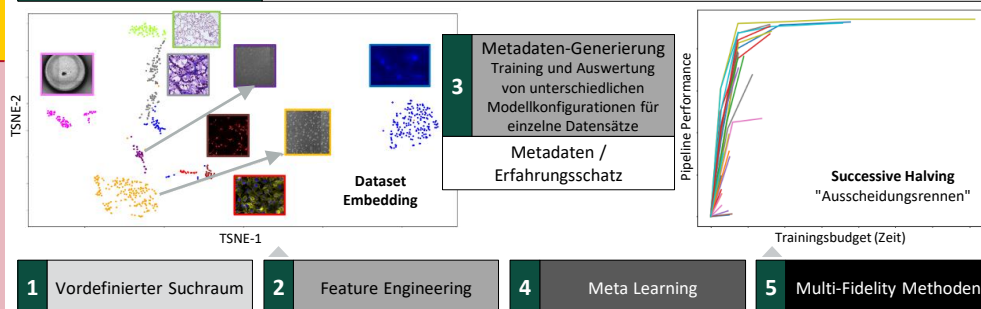
Da die entwickelte Software zur automatisierten Auswertung nicht auf nur einen speziellen Anwendungsfall zugeschnitten ist, profitieren die Anbieter zahlreicher Anwendungsfälle und -nischen von den Ergebnissen. Biotechnische und medizinische Labore können mit der Software schneller belastbare und vergleichbare Ergebnisse erzielen. Software-Anbieter und KI-Entwicklungsunternehmen können ihr Produktportfolio um vergleichbare AutoML-Module erweitern und diese auf andere Anwendungsfelder transferieren, z. B. auf die Erkennung von Oberflächendefekten in der Produktionstechnik. Hersteller optischer Geräte, Laborequipment und photonischer Komponenten für die Zell-, Stammzellen- und Tumorforschung können durch Integration der Ergebnisse in bestehende Lösungen den Zeitaufwand für die Automatisierung der Auswertung reduzieren und deren Genauigkeit erhöhen. Unternehmen der genannten, KMU-dominierten Branchen können mit den Ergebnissen neue Geschäftsmodelle erschließen, z. B. den Vertrieb Hardware-unabhängiger Software-Abonnements.

Umsetzung

Die entwickelte Software als Ganzes steht Industrie, Kliniken und Wissenschaft für die Verwendung zur Verfügung. Die einzelnen Teilmodule sind für separate Verwendungen einsetzbar.

Ansatz des domänen-spezifischen AutoML-Systems

Sukzessive Eingrenzung des Suchraums bis die performanteste Deep Learning Pipeline identifiziert ist



Funktionsweise des domänenspezifischen AutoML-Systems. In fünf Schritten wird der Suchraum sukzessive eingegrenzt, bis die performanteste Deep Learning Pipeline identifiziert ist.