

Projektplan

AIxCell: Cell Culture Analysis Tool (21361 N)

Forschungsziel

Die Analyse biomedizinischer Bilddaten durch Experten, z. B. in der Zellmikroskopie, ist personal- und kostenintensiv sowie anfällig für anwendungsspezifische Fehler. Deep Learning (DL) ermöglicht zwar Hardware-unabhängig automatisierte und objektive Bildauswertungen, wird aber bisher aufgrund des hohen Aufwands für Auswahl und Konfiguration der DL-Modelle nicht eingesetzt. Projektziel ist es ein Software-Tool zur automatisierten Vorverarbeitung der Daten, Algorithmusauswahl und Konfiguration von DL-Modellen zu entwickeln. Dazu dienen selbst lernende Entscheidungslogiken, Hyperparameteroptimierung, DL-Bibliotheken und Vortraining neuronaler Netze, sodass biologische Experten dazu befähigt werden DL-Modelle für die Zellkulturzustandsanalyse zu verwenden.

Arbeitsdiagramm

Arbeitspaket	Zeitraum																							
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
AP 1: Anforderungen & Anwendungsszenarien	■	■	■	■																				
AP 2: Datenintegration			■	■	■	■	■																	
AP 3: Entwicklung der Algorithmenauswahl AutoKonfig				■	■	■	■	■	■	■	■	■	■	■	■	■	■							
AP 4: Entwicklung einer Plug-in DL-Bibliothek				■	■	■	■	■	■	■	■	■	■	■	■	■	■							
AP 5: Zusammenführung der Ergebnisse															■	■	■	■	■	■	■			
AP 6: Bewertung und Validierung																				■	■	■	■	■
AP 7: Verbreitung & Dokumentation			■	■	■	■	■																	
AP 8: Projektmanagement	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Meilensteine				◆ MS1				◆ MS2							◆ MS3				MS4	◆			MS5	◆

MS 1 = Anforderungen sind definiert

MS 2 = Datenpipeline ist implementiert

MS 3 = AutoKonfig und DL-Bibliothek sind implementiert

MS 4 = Softwaredemonstrator funktionsfähig

MS 5 = Demonstrator an Beispielen des PA evaluiert

Lösungsweg zur Erreichung des Forschungsziels

Die Anwendung von DL zur Zellklassifizierung erfordert Expertise im Bereich Machine Learning (ML). Die überwiegende Mehrheit der biologischen Experten hat weder das Fachwissen noch die Zeit diese Aufgabe allein zu bewältigen. Damit die biologischen Experten eine DL-basierte Zellklassifizierung durchführen können, muss die Interaktion mit dem DL intuitiv sein.

Der Anwender lädt annotierte Bilder von adhärennten Zellen, zusätzliche Informationen über den Anwendungsfall (z. B. Anzahl der Zellen, Rundheit, Länge, etc.) sowie Anforderungen und Grenzen in die Software. Das Software-Tool AlxCell verarbeitet die Daten automatisch vor und wählt einen geeigneten DL-Algorithmus inklusive seiner Konfiguration (z. B. Hyperparameter, Anzahl an Layer, etc.) aus. Der bestgeeigneten DL-Algorithmus wird implementiert und dem Benutzer zur Verfügung gestellt. Zusätzliche Informationen über den Auswahlprozess des DL-Algorithmus und die Performance des Modells werden bei Bedarf ebenfalls an den Benutzer weitergegeben. Somit wird die Zeit zwischen der Datenaufnahme und einem geeigneten DL-Algorithmus verkürzt. Dies ermöglicht mehr Anwendern die Vorteile von DL zu nutzen. Dieses Konzept ist in Abbildung 1 dargestellt. Dabei werden sowohl DL-Expertise als auch Expertise aus dem Bereich der Lebenswissenschaften und Biologie im projektbegleitenden Ausschuss zusammengeführt.

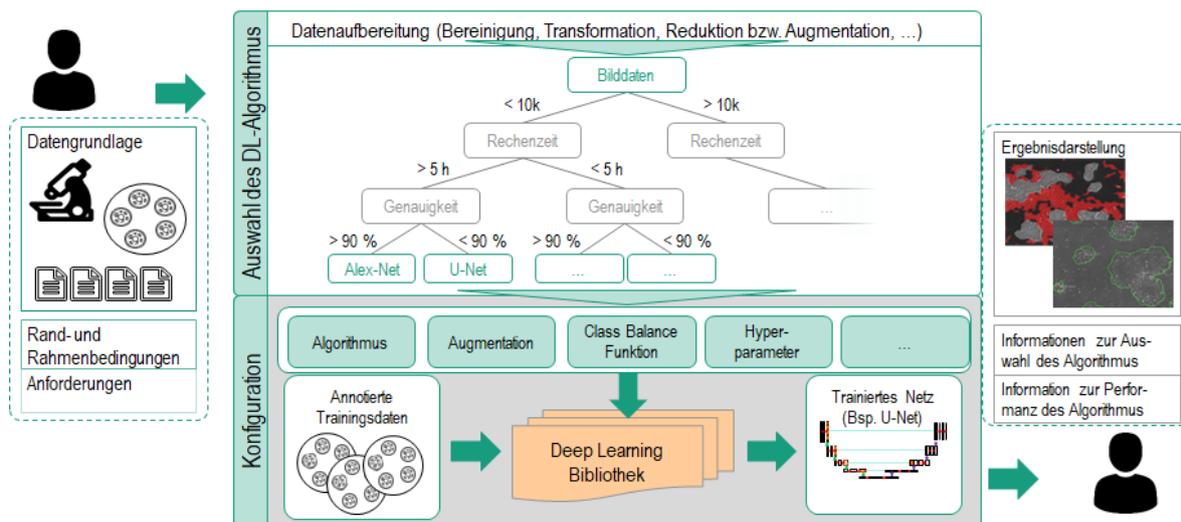


Abbildung 1: Konzept von AlxCell.

Die Aufgabe der automatischen Auswahl des geeigneten DL-Algorithmus und seiner Konfiguration kann in zwei Schritte unterteilt werden:

1. Empfangen und Vorverarbeiten der Eingangsdaten;
2. Auswählen des DL-Algorithmus und seiner Konfiguration.

Die Lernaufgaben in diesem Anwendungsbereich sind Klassifizierung und Clustering. Die Klassifikationsaufgabe ist eine überwachte Lernaufgabe; daher ist es notwendig, dass die Eingangsdaten zumindest teilweise annotiert sind. Wie in der Abbildung 2 dargestellt, interagiert der Benutzer mit einem

Frontend, das es dem Benutzer ermöglicht, seine Daten zu annotieren und hochzuladen, dazu wird ein Dataannotationsguide entwickelt. Bei Auswahl des DL-Algorithmus werden biologische und technische Aspekte wie Randbedingungen, zeitliche Anforderungen oder Genauigkeitsklassen abgefragt.

Die meisten gängigen Datenvorverarbeitungstechniken können automatisch durchgeführt werden. Dennoch ist eine automatische Datenvorverarbeitung nur möglich, wenn der Anwendungsbereich bekannt ist, z. B. klassenweise Konfluenzbestimmung. Für diesen Anwendungsbereich verarbeitet AlxCell die Dateneingabe automatisch vor wählt den bestgeeignetsten Algorithmus aus und trainiert diesen automatisch. Das Model wird dem Nutzer zur Verfügung gestellt. Das Ergebnis, z. B. die Genauigkeit und Geschwindigkeit der Erkennung wird dem Auswahlalgorithmus für zukünftige Lernprozesse zur Verfügung gestellt.

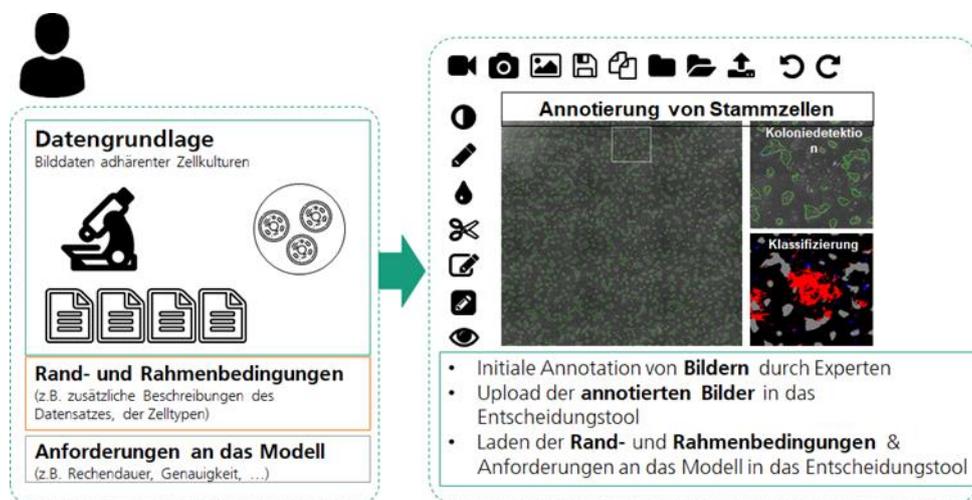


Abbildung 2: Annotation und Import des Datensatzes

Zur Auswahl des geeigneten DL-Algorithmus und seiner Konfiguration wird eine Logik basierend auf Meta-Learning entwickelt (AutoKonfig). Historische Daten (verschiedene Zelltypen, ihre Bilder, Anmerkungen und Beschreibungen) und die Performance verschiedener DL-Algorithmen werden als Trainingsdaten für den Meta-Learning-Algorithmus verwendet. Der Meta-Learning-Algorithmus identifiziert Zusammenhänge zwischen den Eingangsdaten und der Performance des unterschiedlich konfigurierten DL Algorithmus und erstellt ein entsprechendes Modell, z. B. einen Entscheidungsbaum oder ein neuronales Netzwerk (vergleiche Abbildung 1).

AutoKonfig liefert Informationen über den am besten geeigneten DL-Algorithmus und dessen Konfiguration. Um die Aufgabe der automatisierten Erkennung zu ermöglichen, ist eine Implementation und Konfiguration des identifizierten Algorithmus erforderlich. Dazu wird eine DL-Bibliothek entwickelt, die verschiedene Algorithmen als Prototypen in einer „Plug-in“ Struktur bereitstellt und über Schnittstellen zur AutoKonfig verfügt. Des Weiteren wird eine Daten-Pipeline verwendet, um Datensätze hardwareunabhängig hinzuzufügen. Die DL-Bibliothek umfasst folgende Funktionen:

1. Interoperabilität: Schnittstelle zur AutoKonfig zur Konfiguration des Prototyps von DL-Algorithmen , Verwendung von Augmentationstechniken und Feedbackfunktion über das Lernergebnis.
2. Anwendbarkeit: Datenverarbeitungspipeline für schnelles Training und Verarbeitung von zuvor erlernten Modellen
3. Erweiterbarkeit: Plug-in-Struktur für neuer DL-Algorithmen

Für eine Erkennungsaufgabe liefert AutoKonfig zusammen mit den Ergebnissen aus dem Dataannotationguide und der DL-Bibliothek eine oder mehrere gebrauchsfertige Konfigurationen der DL-Modelle. Rechenzeiten und Genauigkeiten werden dem Nutzer für die trainierten Netzwerke angezeigt. Durch das vorherige Trainieren verschiedener Use-Cases wird es möglich sein, AutoKonfig so zu konfigurieren, dass bereits eingelernte Modelle für gleiche Aufgaben wiederverwendet werden können. Dies spart insbesondere Rechenzeit und Aufwand. Abbildung 3 zeigt eine Übersicht über die Konfigurationslogik für die einzelnen DL-Modelle.

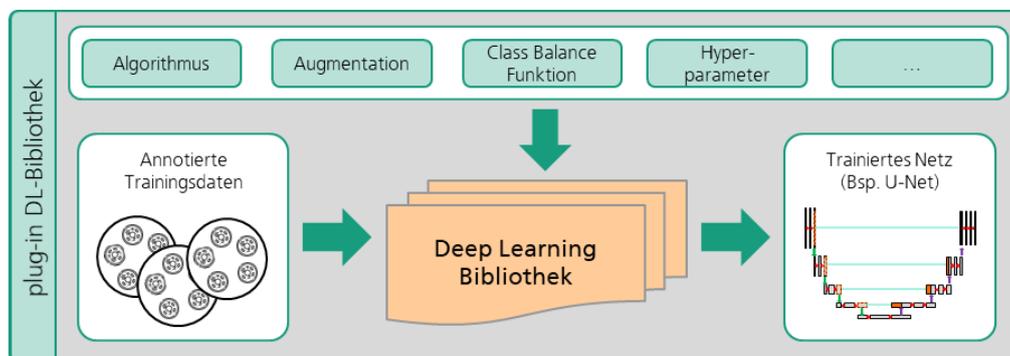


Abbildung 3: Struktur der Plug-In DL-Bibliothek: Die DL-Bibliothek implementiert mehrere DL-Modelle, mit der Schnittstelle AutoKonfig wird eine Konfiguration ausgewählt, das Netzwerk mit den annotierten trainiert und anschließend zur Verfügung gestellt.

Nutzen und wirtschaftliche Bedeutung des Forschungsthemas

Der im Rahmen dieses Projektes vorhergesehene Standardbenutzer ist biologischer Experte, aber auch andere Endnutzer, wie Mediziner, Virologen oder andere Experten aus dem Bereich der Lebenswissenschaften können AIxCell verwenden. Wie in Abbildung 4 dargestellt, beinhaltet die Ergebnispräsentation Ergebnisse der DL-Algorithmen für den Anwendungsfall. Darüber hinaus werden Informationen über die Leistung des Modells bereitgestellt (z. B. F1 Score, Accuracy, Recall, MCC, Interoperator Varianz). Bei Bedarf werden darüber hinaus zusätzliche Informationen bereitgestellt, wie z. B. warum der DL-Algorithmus ausgewählt wurde, wie das DL-Modell konfiguriert ist und welche Alternativen es gibt.

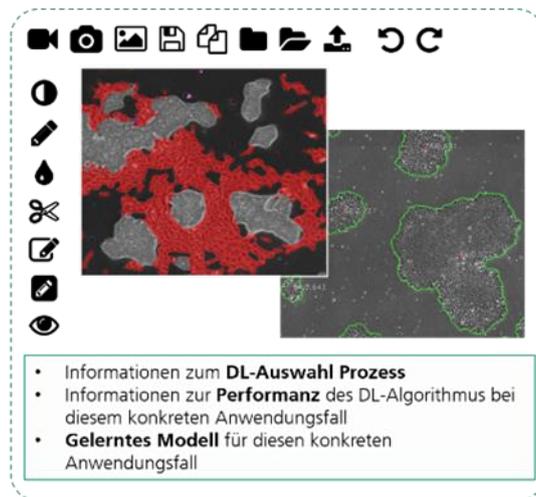


Abbildung 4: Präsentation der Ergebnisse für den Biologischen Experten

Das Softwaretool ermöglicht es dem biologischen Experten, DL in neuen Anwendungsfällen anzuwenden. Die Gesamtlösung der 2D-Bildinterpretation lässt sich auf weitere Anwendungsbereiche übertragen. Nicht nur im Bereich der Biologie, sondern auch in allen Bereichen der Lebenswissenschaften und der Medizin, wie z. B. der medizinischen Bildgebung, der Histologie, der Bestimmung von Viren auf Zellrasen, der Analyse der Gelelektrophorese oder der Funktion von Medikamenten, kann die entwickelte Software als eigenständiges Werkzeug eingesetzt werden. Der Anwender benötigt keine Vorkenntnisse im Bereich DL und ist in der Lage, Bilddaten hardwareunabhängig automatisiert auszuwerten.

Projektbegleitender Ausschuss

Unternehmen
ALS Automated Lab Solutions GmbH ^{KMU}
Bayer AG
Cellmatiq GmbH ^{KMU}
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}
Taorad GmbH ^{KMU}
Sonstige
Ruhruniversität Bochum
Stammzellnetzwerk NRW e. V.
Uniklinik Köln
Uniklinik RWTH Aachen
Universitätsklinikum Bonn
Universitätsmedizin Göttingen

Das IGF-Vorhaben Nr. 21361 N der Forschungsvereinigung Feinmechanik, Optik und Medizintechnik wird über die AiF im Rahmen des Programms zur Förderung der Industriellen Gemeinschaftsforschung (IGF) vom Bundesministerium für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestages gefördert.