

Drücken Sie „Play“, um mit OCR zu beginnen

Nicht KI, sondern wie man KI einsetzt, macht den Unterschied



Die Erwartungen an eine OCR (Optical Character Recognition) sind hoch. Heutzutage erwarten Anwender, dass sie alle Zeichen out-of-the-box erkennt und richtig deutet, ähnlich wie ein Mensch es kann. Darüber hinaus sollte die Genauigkeit konsistent bleiben, unabhängig von Veränderungen in der Lichtsituation oder anderen Umgebungsbedingungen. Regelbasierte Ansätze haben den Nachteil, dass ihnen der Interpretationsspielraum fehlt, der bei Ansätzen mit neuronalen Netzen durch das Training mit vielen verschiedenen Beispielen erworben wird und ihre enorme Leistungsfähigkeit erklärt. Der alleinige Einsatz modernster Technologie ist jedoch nicht ausreichend für den Projekterfolg - vielmehr geht es darum diese einfach und performant nutzbar und benutzerfreundlich wartbar anzubieten. Denn auch bei den auf Deep-Learning basierenden OCR-Systemen ist Qualität, Ausführungsgeschwindigkeit und Benutzerfreundlichkeit alles andere als selbstverständlich.

Optische Zeichenerkennung

Die optische Zeichenerkennung zählt auch heute noch zu einer der

schwierigsten Disziplinen in der Bildverarbeitung und maschinellen Intelligenz. Schon die reine Vielfalt an möglichen Schriftzeichen und Verfahren, wie Zeichen auf verschiedensten Oberflächen angebracht werden, vermittelt eine Vorstellung der Herausforderungen. Die Schwierigkeiten solche komplexen visuellen Daten in klare, strukturierte Texte umzuwandeln, umfassen Schmutz, Reflexionen sowie Formfehler durch

Ritzen, Prägungen oder Lasergravuren auf festen Materialien. Zudem können überlagerte oder unvollständige Zeichen, sowie eine generell niedrige Pixelauflösung der Bild-daten dazu führen, dass sich Zeichen rasch kaum mehr voneinander unterscheiden lassen. So wird z. B. eine 8 schnell zu einer 3. Der Bild-verarbeitungsmarkt entwickelt sich kontinuierlich weiter, um die Genauigkeit und Zuverlässigkeit der



Bild 1: Die DENKnet OCR liest in vielen Anwendungsfällen auch ohne Feintuning schon sehr sicher, beispielsweise Reifennummern mit wenig Kontrast.

Autor:
Dipl.- Ing. Heiko Seitz
Technischer Autor
IDS Imaging Development
Systems GmbH
www.ids-imaging.de



Bild 2: Stark verformte und kleine Nummern auf Kronkorken werden sicher gelesen

Texterkennung zu verbessern. Doch was sind die entscheidenden Faktoren für die Auswahl eines OCR-Systems?

Vielbelesene Basis mit reproduzierbarer Genauigkeit

Eine OCR muss von Anfang an einfach funktionieren und eine hohe Leseleistung bieten, um zu überzeugen. Dazu bedarf es einer gut entwickelten Netzarchitektur, die mit vielen variantenreichen Trainingsbildern vortrainiert wurde. Hier sind Situationen aus realer Anwendung genauso unverzichtbar wie der Einsatz synthetischer Daten. Damit können nicht nur viele zusätzliche Sonderfälle und Variationen gelernt werden, das sorgt auch für eine weit aus robustere Erkennung der relevanten Merkmale. Denn gerade in der industriellen Automation darf nichts dem Zufall überlassen sein.

AI Vision-Lösung für individuelle Bildanalysen

An dieser Stelle setzt DENKnet an, die AI Vision-Lösung für individuelle Bildanalysen. Neben führender KI-Technologie, steht Anwendern hier ein enorm performantes und ständig weiterentwickeltes OCR-Modell zur Verfügung. Alle Entwicklungsschritte sind dabei streng versioniert, sodass Anwendungsentwicklungen auf definierte Versionen zurückgreifen können, aber auch die Möglichkeit haben, auf eine neue verbesserte Version zu aktualisieren, um stets vielseitig und robust zu lesen. Zur Qualitätssicherung kann die Performance und Reproduzierbar-

keit der trainierten Netze in einem Quality Center gegen Beispieldatensätze geprüft und verifiziert werden, bevor eine Produktionsanlage mit neuer Software aktualisiert wird.

Von Transformern und Large Language Models

Eine weitere positive Eigenschaft eines guten OCR Modells liegt in der Fähigkeit, nicht nur einzelne Zeichen, sondern die Zusammenhänge – bei Zeichenfolgen, wie bspw. Seriennummern oder Worten – zu kennen und dieses Wissen bei der Zeichenerkennung zu berücksichtigen. Je besser die OCR auch Folgezeichen vorhersagen und das Leseergebnis damit gewichten kann, desto robuster und präziser können spezielle Anwendungsfälle damit gelöst werden. Die generativen und kombinatorischen Eigenschaften von Transformer-Netzen oder Large Language Modellen (LLM), wie sie in ChatGPT verwendet werden, könn(t)en solche Vorhersagen und damit auf die Lesequalität nochmal positiv beeinflussen. Doch dabei sollte man bedenken, dass diese Architekturen in der Ausführung eher langsam sind und sehr viele Systemressourcen benötigen.

Cutting-Edge-Technologien

Umso wichtiger, dass der Einsatz solcher Cutting-Edge-Technologien im richtigen Maß erfolgt, um die Anforderungen aus Kunden-Use-Cases optimal zu unterstützen. Und gerade im Automatisierungsbereich sollte sich eine Bildverarbeitung nicht im Sekunden-, sondern eher im niedrigen Millisekunden-

Bereich bewegen. Ein trainiertes neuronales Netz sollte deshalb schnell und leichtgewichtig bleiben, um es auf „normaler“ Hardware ausführen zu können. Sind hohe Erkennungsgenauigkeit und Geschwindigkeit im produktiven Einsatz nur mit schier unendlicher Systemleistung möglich, wären Anwendungen damit kaum wirtschaftlich lösbar.

„Der Trend geht dahin die KI kleiner und damit schneller und kostengünstiger in der Ausführung zu machen.“ Daniel Routschka, Sales Manager Artificial Intelligence bei IDS.

Einfaches Korrigieren und Nachtrainieren

Sollte die OCR dann doch mal Zeichen nicht lesen, egal ob der Grund ein Fehler oder ein unbekanntes Zeichen bzw. Schriftart oder Sprache war, ist es wichtig, dass der Anwender mit wenig Aufwand das Leseergebnis korrigieren oder auch beliebige neue Zeichen trainieren kann. Doch bei diesem Feintuning handelt es sich nicht um ein einfaches „Weiterlernen“ des Netzes. Man stelle sich vor, dass das OCR-Modell beispielsweise mit 2 Millionen Bildern trainiert wurde und der Benutzer dem OCR-Modell nun mit einigen wenigen eigenen Bildern etwas Neues beibringen möchte. Mit welcher Gewichtung geht eine

solche Info in das Modell ein, um sowohl etwas zu bewirken, dabei aber auch nicht alles zu verändern?

KI sinnvoll erweitern

Und genau hier ist viel Know-how des Anbieters gefragt, die KI so zu erweitern, dass durch eine derartige Anpassung nicht bisherige stabile Erkennungen negativ beeinflusst werden. Ein Beispiel: Eine OCR hat aus irgendeinem Grund Probleme mit Zahlen und der Anwender annotiert im Trainingsprozess nur Zahlen, nie Buchstaben. Dabei gilt es durch eine intelligente „Wissenssicherung“ zu verhindern, dass dieses Netz irgendwann nur noch Zahlen erfolgreich lesen kann, weil es denkt, es müsse keine Buchstaben lesen.

Der DENK Vision AI Hub generiert deshalb beim Feintuning der DENKnet OCR für alle neuen Bild-daten passende künstliche Daten, um das Netz im richtigen Maß weiter zu trainieren und zu gewichten. Das verhindert, dass die OCR, egal wie lange sie weitertrainiert wird, seine bisherigen Fähigkeiten verliert. Dabei bleibt das „Nachtrainieren“ für den Benutzer des Vision AI Hubs einfach in der Handhabung und durch das Cloud-basierte Training im Hintergrund schnell und performant. Im besten Fall sind

DENKnet OCR – Das macht den Unterschied

- **Synthetische Daten:** Bei jedem Upload neuer Bilder werden automatisch Bildvarianten erzeugt, um die Modell-Fähigkeiten zielgerichtet zu erweitern und zu stabilisieren.
- **Benutzerfreundlichkeit + Zeitersparnis:** Intuitive Werkzeuge, wie „Autoprediction“ und „1-Click Annotation“ erfordern kein Vorwissen und verkürzen Test-, Vorbereitungs- und Wartungszeit.
- **Cutting-Edge-Technologie:** Erkenntnisse der neuesten Netzwerkkonstrukturen, wie Transformer oder Large Language Models, fließen kontinuierlich in die Entwicklung der DENKnet OCR ein.
- **Smart Architecture:** Vollautomatisches Training wählt selbständig die am besten geeignete Architektur für die Aufgabe
- **Cloud Training:** Immer Up-to-date mit Cutting Edge Technologie und kontinuierlicher Verbesserung der Netzbasis
- **Lokale Ausführung schnell und wirtschaftlich:** Ziel ist ein optimal arbeitendes genaues und auch gleichzeitig schlankes und schnelles Modell für die lokale Ausführung in einer geschlossenen Anwendungsumgebung



Bild 3: Auch Infos auf Trennscheiben mit erheblicher Überdrückung bei stark inhomogenem Hintergrund stellen kein Problem dar.

Gefahr dass unterschiedliche Bildsysteme oder Software-Versionen zu unterschiedlichen Ergebnissen führen, können bspw. Änderungen an der Netzarchitektur vorgenommen oder die Erzeugung synthetischer Zusatzdaten optimiert werden. Das geht so im direkten Austausch, ohne Zeitverlust direkt im Kunden-Use Case. Der Verzicht auf den Versand sensibler Daten minimiert zudem das Risiko eines unbefugten Zugriffs.

OCR einfach und wirtschaftlich aus einer Hand

Im AI Vision-Umfeld tummeln sich viele Anbieter von OCR-Lösungen und es gibt ein regelrechtes Wettrennen um die besten Netze. Für versierte Anwender sind zudem viele Open Source Werkzeuge und öffentlich zugängliche Netzarchitekturen verfügbar, mit denen man schnell erste Erfahrungen sammeln und Ergebnisse erzielen kann. Doch ohne tieftechnisches Wissen, wie sich KI-Technologie bzw. Cutting Edge Netzwerke und Large Vision Modelle, wirtschaftlich und performant einsetzen und kombinieren lassen, bleiben viele OCR-Aufgaben ungelöst.

Anders beim Industriekamerahersteller IDS: Zusammen mit der AI Vision-Lösung DENKnet können alle Bildverarbeitungs-komponenten für schnelle, zuverlässige und wirtschaftliche OCR-Aufgaben einer Hand geliefert werden. Kunden profitieren, weil es funktioniert. Und ausprobieren kostet nichts. „Just Press Play“ ◀

die Basis-Fähigkeiten der OCR so gut, dass Anwender gar nicht mehr nachtrainieren müssen.

Vorteil Cloud-Training

Alle Funktionen und Dienste des DENK Vision AI Hub basieren vollständig auf Cloud-Technologie. Dadurch findet das Feintuning auf eigenen Bilddaten auf einer stets aktuellen und kontrollierten Software-Basis statt und nicht auf irgendeiner Software-Version auf irgendeinem lokalen Hard-

waresystem. Das dort verwendbare OCR-Modell wird durch die kontinuierliche Weiterentwicklung im technischen Backend immer resistenter gegenüber bereits gelösten Schwierigkeiten. Dadurch können immer mehr Kundenanwendungen sogar ohne größere Anpassungen oder ein Nachtraining verwendet werden. „Press Play“ ist wie ein Auftrag an die DENKcloud, die im Hintergrund eine Vielzahl passender Netzmodelle mit unterschiedlichen Architekturen trainiert und dem Anwen-

der letztendlich das beste Ergebnis zur Verfügung stellt.

Supportfall

Auch im Supportfall ist die Cloud-Lösung ein Mehrwert für den Anwender. Sollte es mit Daten eines Use-Cases Schwierigkeiten geben, z. B. bei unbekannte Schriftzeichen, kann technische Unterstützung im Backend schnell Abhilfe schaffen und Erkennungsleistung positiv beeinflussen. Ohne Daten exportieren/importieren zu müssen und ohne die

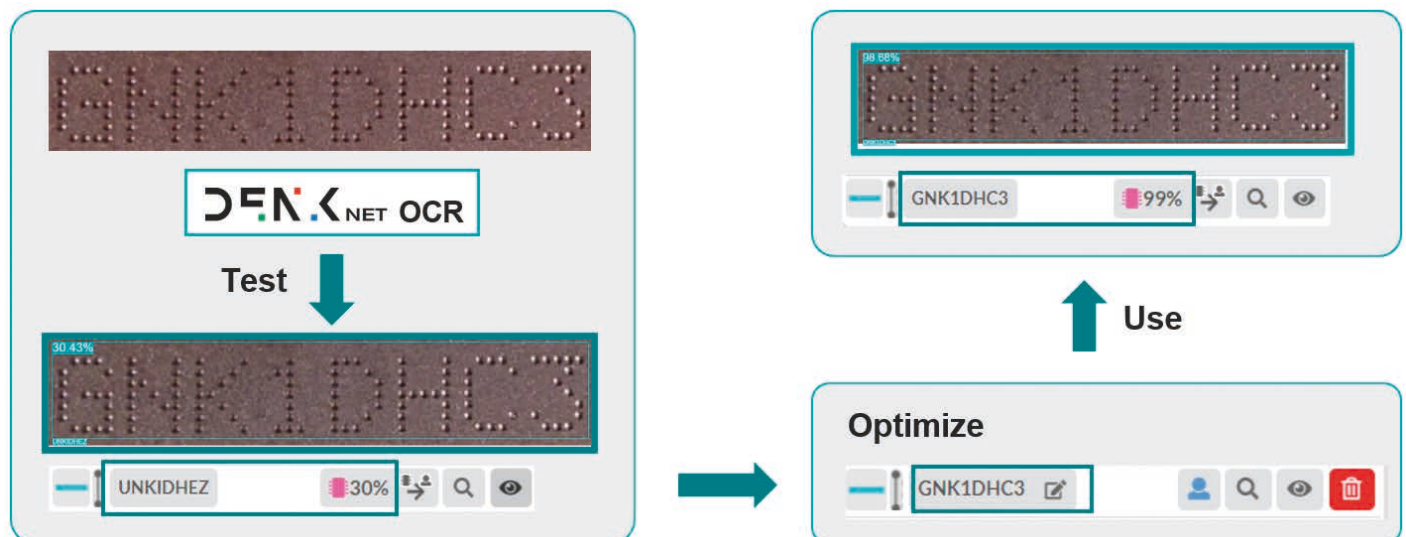


Bild 4: Das Feintuning der DENKnet OCR im DENK Vision AI Hub erfordert nur wenig Anwenderinteraktion und verbessert sehr schnell die Lesequalität