

## Fallstudie: KI in 20 Tagen

**Abstrakt: KI Technologie in der Bilderkennung ist heute bereit für die produktive Anwendung. Bei den Einsatzmöglichkeiten und der konkreten Umsetzung von KI-Lösungen sind aber heute noch viele Fragen offen. Mit unserer Fallstudie zeigen wir das Ergebnis einer konkreten KI Umsetzung, welche auf die überwiegende Anzahl von Anwendungsfällen übertragbar ist. Der Gesamtaufwand für die Fallstudie wurde auf 20 Arbeitstage beschränkt.**

Als Komplettlösungsanbieter für kundenspezifische Deep-Learning-Lösungen bietet die Evotegra GmbH das gesamte Spektrum von Beratung und Datenservices bis hin zur Prozessintegration an. Mit der Fähigkeit, kostengünstige Lösungen in großem Maßstab anzubieten, richten wir uns an kleine und mittlere Unternehmen, die keine Möglichkeit haben ihre eigene KI-Kompetenz aufzubauen.

### Was ist künstliche Intelligenz?

Die heute in der Praxis verwendete „schwache künstliche Intelligenz“ ist eine sich selbst optimierende mathematische Annäherung an eine unbekannte komplexe Funktion. Die Optimierung erfolgt anhand von Beobachtungen, die der künstlichen Intelligenz in Form von Daten während der Trainingsphase zur Verfügung gestellt werden. Während der anschließenden Anwendungsphase lernt eine schwache KI nicht mehr dazu.

Da heutige künstliche Intelligenz zu keinem Zeitpunkt ein Bewusstsein über das Erkannte entwickelt, ist KI dabei vollkommen objektiv.

### Die Vor- und Nachteile künstlicher Intelligenz

Die Vorteile von KI in der Objekterkennung liegen in der Verbindung von Genauigkeit und Geschwindigkeit in Kombination mit der Flexibilität praktisch alles zu erkennen, was ein Mensch sehen kann.

Die Vorteile von KI:

- Hohe Erkennungsgenauigkeit
- Hohe Geschwindigkeit
- Robuste Erkennung
- Hohe Anpassungsfähigkeit (Lernfähigkeit)

Die Nachteile:

- benötigt große Datenmengen
- benötigt Zeit zum trainieren (1-2 Tage)

Folgende Objekteigenschaften eignen sich für die Erkennung mit KI :

- Hohe Varianz

- undefinierte räumliche Ausrichtung
- undefinierte Lichtbedingungen
- begrenzte Anzahl zu erkennender Objekte (<500)
- geringe Änderung der Menge zu erkennender Objekte

z.B. Verkehrsschilder, Gefahrgutkennzeichen, Personen, Fahrzeuge, Flaschen...

Eine Kontraindikation ist gegeben, wenn sich die Menge zu erkennender Objekte häufig ändert oder eine große Anzahl (>1000) von unterschiedlichen Objekten erkannt werden muss.

### Die Anwendung von künstlicher Intelligenz

Während KI mittlerweile eine höhere Erkennungsgenauigkeit erreichen kann als der Mensch, sind KI-Lösungen nicht grundsätzlich verlustfrei übertragbar. Jeder Unterschied in der Anwendung im Vergleich zu den Trainingsszenarien, kann die Erkennungsrate negativ beeinflussen. Auf der anderen Seite erhöht das Hinzufügen weiterer Szenarien die Wahrscheinlichkeit falscher Erkennungen. Während dies für Verbraucherprodukte akzeptabel sein mag, ist diese Eigenschaft für industrielle Lösungen in der Regel nicht geeignet. Daher glauben wir nicht an klassische Produkte mit integrierter KI (eine KI für alle Anwendungsfälle), sondern an einzelne plattformbasierte Lösungen, die für ein spezielles Szenario optimiert sind.

Die heutige KI-Technologie ist mittlerweile für die Produktion geeignet. Neuronale Netzwerke können zuverlässig trainiert und ausgeführt werden, C++ und Netzwerkoptimierung ermöglichen die Prozessintegration. Eine breite Palette verfügbarer Hardware, von eingebetteten Systemen bis hin zu High-End-Rechenzentrumslösungen, ermöglicht den Einsatz von KI-Lösungen in nahezu jedem Szenario.

Der Einsatz von KI-Lösungen in der Nähe von Maschinen (Edge-Processing) ist für Industrie- und Bildverarbeitungslösungen besonders wichtig. Während die hohe Bandbreite von Kameradaten in der Cloud nicht effizient verarbeitet werden kann, verhindern auch Unternehmensrichtlinien unter Umständen eine cloud-basierte Verarbeitung sensibler Produktionsdaten.

### Daten

Eine KI-Lösung ist niemals besser als die Qualität der verwendeten Daten. Für ein allgemeines Erkennungsproblem mit 100 verschiedenen Klassen empfehlen wir je nach Szenario insgesamt zwischen 100.000 und 1.000.000 Datensätze für das Training der KI. Die notwendigen Daten können jedoch inzwischen in hochautomatisierten Verfahren in großen Mengen und kurzer Zeit in hoher Qualität zu geringen Kosten gewonnen werden.

Darüber hinaus ist es wichtig Änderungen ohne große Zusatzaufwände durchführen zu können und die Datenqualität während der gesamten Projektlaufzeit sicher zu stellen.

## Aufbau der Fallstudie

Der zentrale Vorteil künstlicher Intelligenz ist ihre Lernfähigkeit und somit ihre Anpassungsfähigkeit an beliebige Aufgabenstellungen. Um die Übertragbarkeit auf eine maximale Anzahl potentieller Anwendungsfälle in der Industrie oder im Verkehr zu gewährleisten, wurde eine besonders anspruchsvolle Aufgabenstellung gewählt.

Die Anforderungen an die Objekterkennung sind:

- Kleine Objekte bzw. große Entfernung zur Kamera
- Unkontrollierte Lichtbedingungen
- Komplexer Hintergrund
- Dynamische Szenen
- Bis zu 50 m/s Relativgeschwindigkeit von Objekt zu Kamera
- Mobiler Aufbau
- Maximal 30W Stromverbrauch
- Keine Cloud
- Daten aus dem öffentlichen Raum

Die überwiegende Anzahl von potentiellen Anwendungsfällen hat in der Regel geringere Anforderungen.

Als Ziel für die Objekterkennung wurden daher deutsche Verkehrsschilder festgelegt. Die Fallstudie geht von der Annahme aus, dass keine Daten vorhanden sind. Die Daten müssen daher im Rahmen der Fallstudie erfasst und für das Training der Künstlichen Intelligenz vorbereitet werden. Um einen qualitativ hochwertigen Datensatz mit hoher Diversität zu erhalten wird die Anzahl von Bildern pro Fahrt und Schild auf 5 begrenzt.

Der Gesamtaufwand zur Durchführung der Fallstudie wird auf 20 Personentage begrenzt.

## Festlegung der zu erfassenden Klassen

Der Umfang für die Erkennung der Verkehrsschilder ist der Gesamtkatalog der Straßenverkehrsordnung inklusive komplexer Wegweiser für Autobahn und Bundesstrassen, digitaler Verkehrsschilder und verschiedener Arten von Lichtsignalanlagen. Hierfür wurde eine Klassenbeschreibung festgelegt und dokumentiert. Insgesamt sind im Katalog 297 unterschiedliche Schilder und Markierungen enthalten.

Bsp.:



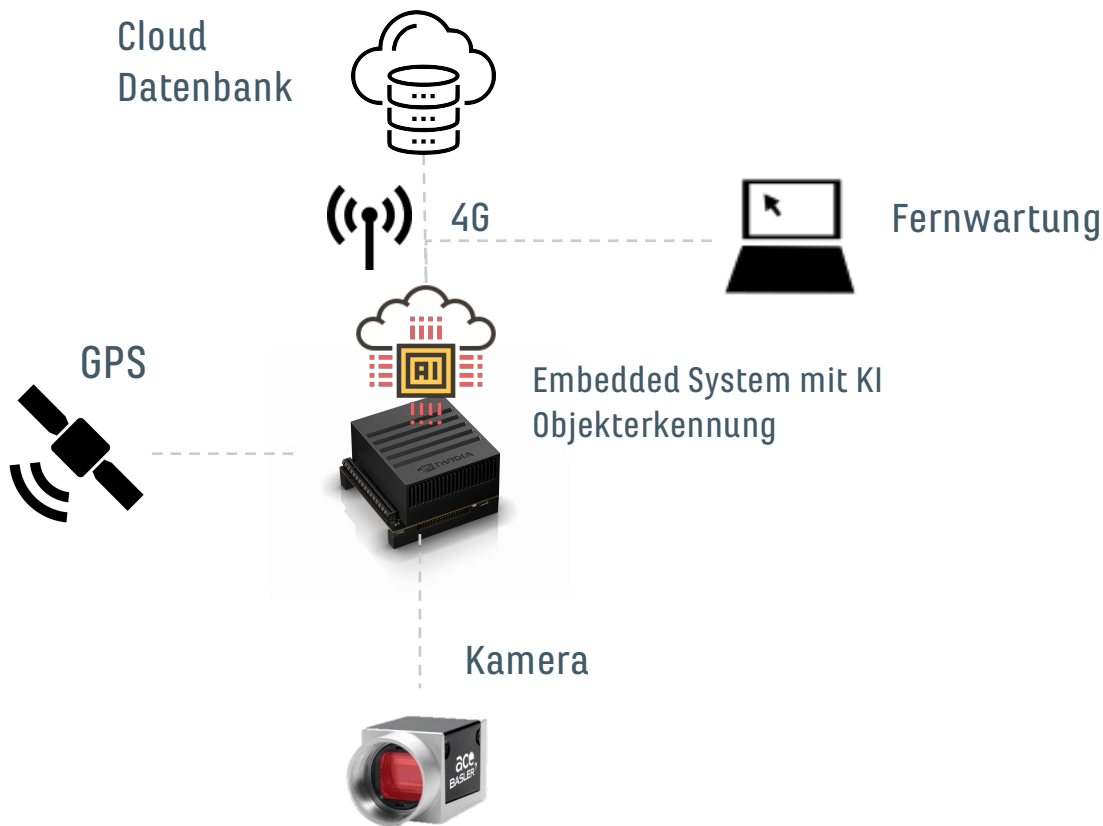
Aufwand: 2 Personentage

### Datenerfassung

Ziel der Fallstudie ist es die Daten für die Verkehrszeichenerkennung vollständig im Rahmen der Fallstudie zu erfassen. Hierfür wird das vollautomatische Datenerfassungssystem der EvoTegra GmbH verwendet. Aufgrund der hohen Relativgeschwindigkeit des Fahrzeug zum Objekt mit bis zu 180 km/h erfolgt die Erfassung mit mindestens 30 Bildern pro Sekunde. Optional kann das System neben den Bilddaten auch räumliche Daten mit Hilfe eines Stereokamerasystems erfassen. Durch den Einbau des Systems in einem handelsüblichen Fahrzeug wird die Nutzung oder Sicherheit nicht eingeschränkt und ein rückstandsfreier Rückbau ist innerhalb weniger Minuten möglich.

Das System besteht aus folgenden Komponenten:

- Embedded System NVIDIA Xavier AGX
- USB-Industriekamera mit Auflösung 1920x1200 und 40hz
- Highspeed LTE Router mit 100 GB monatlichem Volumen
- GPS
- Inverter bzw. KFZ Ladegerät
- Cloud Webservice
- Cloud SQL-Datenbank



Auf dem Embeddedsystem wurde dabei ein vortrainiertes Netzwerk zur Verkehrszeichenerkennung verwendet. Für einige Anwendungsfälle stehen entweder vortrainierte Netzwerke oder entsprechende öffentliche Datensätze bereit. Sind keine vortrainierten Netzwerke oder Daten vorhanden, fällt ein Zusatzaufwand von 5-10 Tagen an.

Mit Hilfe des vortrainierten Netzwerks wird die überwiegende Anzahl von Kamerabildern ohne Schild bereits im Fahrzeug gefiltert. Um neuronale Netzwerke in Echtzeit im Fahrzeug ausführen zu können, wird eine Netzwerkoptimierung verwendet. Bilder mit potentiellen Schildern werden komprimiert, verschlüsselt und zusätzlich mit einer Prüfsumme gesichert. Im Ergebnis wird die Datenmenge so stark reduziert, dass eine Übertragung per LTE aus dem Fahrzeug an einen Cloud-Server erfolgen kann.

Durch die zielgerichtete Datenerfassung, starke Verschlüsselung, Integritätsprüfungen und Identifikation der Kommunikationspartner wird ein verantwortungsvoller Umgang mit den Daten gewährleistet und die Anforderungen der DSGVO abgedeckt. Eine zusätzliche Verbesserung ist die Entfernung personenbezogener Merkmale bereits im Fahrzeug, die wir bereits in der nächsten Version des Datenerfassungssystems integrieren werden.

Da das Fahrzeug lediglich im Rahmen der normalen Nutzung verwendet wurde und das System vollautomatisch arbeitet, wurde für die Datenerfassung im Rahmen der Studie kein Aufwand berechnet. Die Dauer der Datenerfassung hängt daher von der Anzahl der Fahrzeuge und deren Nutzung ab. Im Rahmen der Studie wurden die Daten in einem Fahrzeug über eine Dauer von 2 Monaten mit einer Strecke von ca. 5000km in unterschiedlichen Regionen zu unterschiedlichen Tageszeiten inklusive Nachtfahrten erfasst.

Erfasst wurden 64068 Bilder mit einer Auflösung von 1920x1200 in einem Volumen von 69 GB.

Aufwand: 1 Personentag

### Die Datenverarbeitung

Für das Training eines KI-Systems benötigt man Bilder sowie die Information, wo sich ein Objekt innerhalb des Bildes befindet. Dieser Beschriftungsprozess wird normalerweise manuell oder von Dienstleistern durchgeführt und ist somit in der Regel sehr zeit- bzw. kostenintensiv. Im Rahmen der Fallstudie wurden jedoch hochautomatisierte Verfahren zur Datengewinnung verwendet. Im Vergleich zu einer rein manuellen Bearbeitung kann die Produktivität einer Person so um bis zu 4000% erhöht werden. Neben der Quantität wird auch gleichzeitig die Qualität der Daten deutlich verbessert. Durch die Verwendung von hochautomatisierten Verfahren können fehlerhafte und unvollständige Annahmen identifiziert und korrigiert werden.

Im Rahmen der Datenanalyse zeigte sich z.B. dass verschiedene Arten von Leitbaken existieren. Diese wurden daher in der Definition der zu erkennenden Klassen ergänzt.



Leitbake (605)



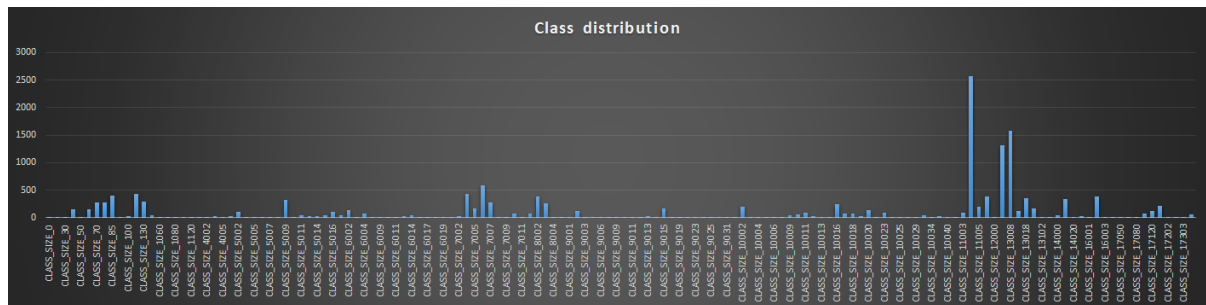
Leitbake (605) Variante 1



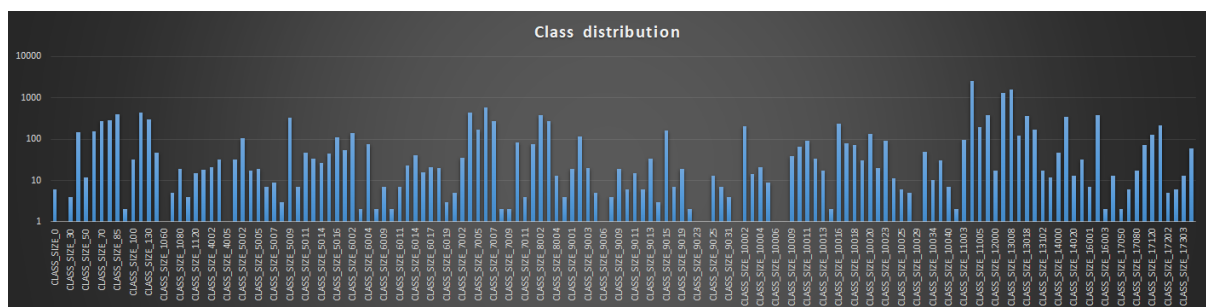
Leitbake (?) Variante 2

Grundsätzlich muss davon ausgegangen werden, dass die initialen Annahmen im Laufe eines KI Projekts mehrfach ergänzt und angepasst werden müssen.

Insgesamt wurden aus den 64068 Bildern 202940 Datenbeschriftungen in 172 Kategorien gewonnen.



[Klassenverteilung]



[Logarithmische Klassenverteilung]

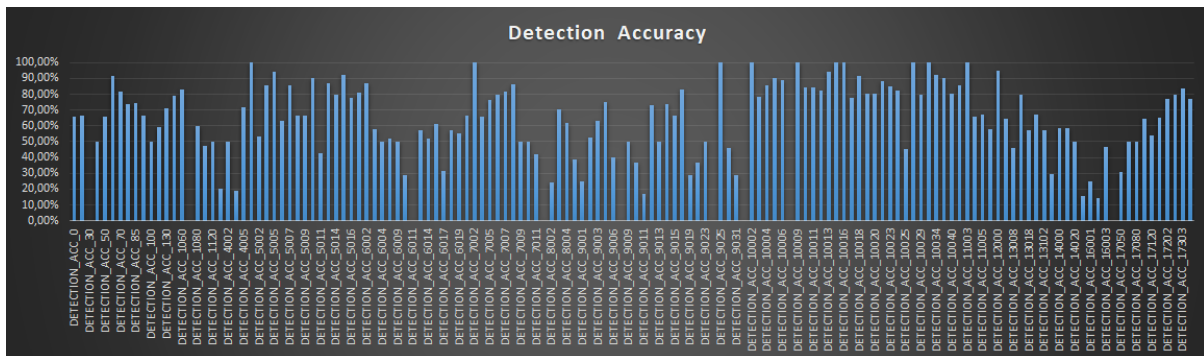
In einem anspruchsvollen Szenario wie der Verkehrsschilderkennung empfehlen wir ca. 10.000 Bilder pro Kategorie zum Training zu verwenden. In einem einfachen Anwendungsfall sollten es mindestens 1000 pro Kategorie sein. Anhand der Klassenverteilung lässt sich erkennen, dass im Rahmen der Studie keine Klasse die empfohlene Marke erreicht hat.

Aufwand: 16 Personentage

## Training

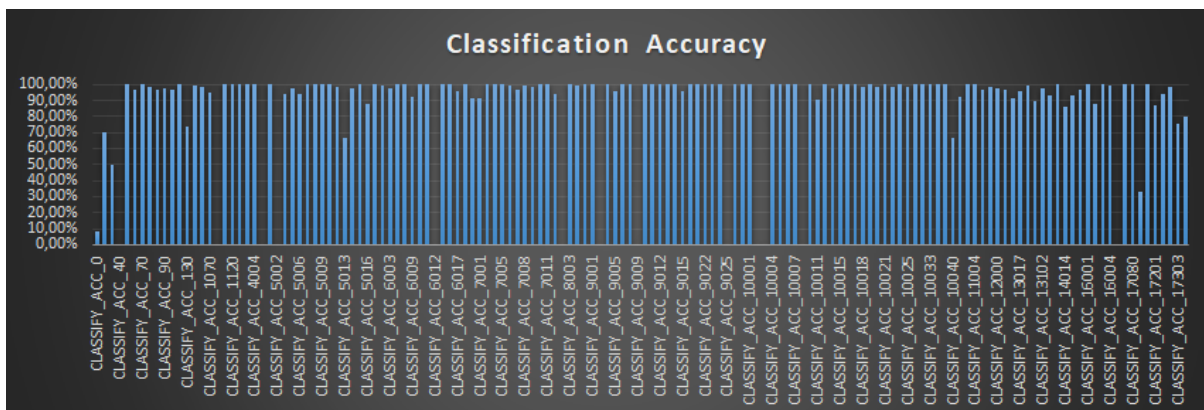
Da die Verteilung der Daten wie zu erwarten sehr ungleich verteilt ist konnten nur 172 Klassen für das Training verwendet werden. Einige Klassen mussten auch zusammengefasst werden.

Nach der Aufteilung in Trainings- und Testdaten wurden rund 20 verschiedene Netzwerke für Objekterkennung und Klassifikation mit unterschiedlichen Parametern trainiert. Die Dauer für einen Trainingslauf beträgt in der Regel 1-2 Tage. Training und Test sind hierbei ebenfalls hochautomatisiert und können parallelisiert werden.



(Erkennungsgenauigkeit nach Klassen)

Die durchschnittliche Erkennungsgenauigkeit erreicht einen F1-Score von 0,7. Das F-Maß kombiniert Genauigkeit (precision) und Trefferquote (recall) durch einen Mittelwert. Die Genauigkeit definiert das Verhältnis von erkannten Schildern zu allen erkannten Objekten und die Trefferquote definiert das Verhältnis von erkannten Schildern zu den tatsächlich vorhandenen Schildern.



(Klassifikationsgenauigkeit nach Klassen)

Die durchschnittliche Klassifikationsgenauigkeit liegt bei 92%.

Insbesondere die Erkennungsgenauigkeit richtet sich hier, wie in den meisten Fällen überwiegend nach der Datenverteilung.

Aufwand: 2 Personentage

## Optimierung und Einsatz

Um Netzwerke in Echtzeit auf dem Embeddedsystem ausführen zu können, müssen diese vorher optimiert werden. Durch die Zusammenführung von verschiedene Netzwerkebenen und der Vereinfachung von Operationen kann eine Beschleunigung der Ausführung um bis zu 500% erreicht werden. Für die Erkennung von kleinen bzw. weit entfernten Objekten wie Verkehrsschildern ist eine hohe Auflösung bei der Ausführung erforderlich.



Für die Ausführung der Netzwerke kommt die Evotegra KI Plattform zum Einsatz. Mit Hilfe der modularen Plattform können unterschiedliche Netzwerkarten aus unterschiedlichen Deep Learning Frameworks ausgeführt werden. Die Plattform integriert bereits unterschiedliche Methoden zur Objektverfolgung und Klassifikation und verfügt über Cloud-, Daten- und Geräteschnittstellen. Die vollständig in C++ entwickelte Plattform kann praktisch in jede Umgebung integriert werden.

Bei einer Auflösung von 960x600 Pixeln können 40-50 Bilder pro Sekunde auf dem Embeddedsystem analysiert werden.

Aufwand: 1 Personentag

## Ergebnisse

Trotz der begrenzten Menge an Daten ist die Erkennung der meisten Klassen bereits gut bis zufriedenstellend. Die Erkennung funktioniert auch Nachts zuverlässig und ist somit robust gegenüber Lichteinflüssen und Bildrauschen. Die Ausführungsgeschwindigkeit in Kombination mit der Auflösung ermöglicht eine zuverlässige Erkennung im Fahrzeug auch bei höchster Geschwindigkeit.

Da eine Verbesserung der Ergebnisse im wesentlichen über zusätzliche Daten erfolgt, lässt sich in KI- Projekten der erfolgreiche Abschluss eines Projekts bereits früh abschätzen. Durch ein Änderungsmanagement und die Qualitätskontrolle von Daten lässt sich das Projektrisiko so minimieren. Zusätzliche Daten können durch vollautomatische Systeme zur Datenerfassung zielgerichtet und zu minimalen Kosten erfasst werden.

Mit den vorgestellten Prozessen können Datensätze in der Größenordnung von Millionen in verhältnismäßig kurzer Zeit erreicht und verwaltet werden. Mit hochautomatisierten Prozessen und flexiblen Plattformen können kundenspezifische KI-Lösungen zu Kosten entwickelt werden, die erstmals hochoptimierte Einzellösungen ermöglichen.

Grundsätzlich ist menschliche Intuition auch mittelfristig nicht durch KI zu ersetzen. Tatsächlich ergänzen sich die menschliche Intuition und die Objektivität einer KI perfekt. Daher sehen wir die optimale Nutzung von KI in der Verbindung der beiden Eigenschaften. KI bietet so dem Menschen eine anspruchsvolle Tätigkeit, die es ihm ermöglicht seine Stärken optimal zu nutzen. Gleichzeitig lässt sich Qualität besser objektivieren und die Produktivität der Mitarbeiter um bis zu 4000% steigern.

KI bietet sich daher heute für eine Vielzahl von visuellen Inspektionsaufgaben an, z.B. bei der Vollständigkeitskontrolle oder Qualitätssicherung. Kameras lassen sich durch individuelle KI Lösungen erstmals als universelle Sensoren wie z.B. bei der visuellen Überwachung von Instrumenten, Schaltern und Prozessen einsetzen. Um die Persönlichkeitsrechte der Mitarbeiter zu

garantieren, kann die Auswertung der Bilder dabei direkt an der Maschine erfolgen. Dadurch müssen Bilder weder gespeichert noch übertragen werden.

