



## Optische Qualitätsprüfung und Prozessoptimierung mittels Deep Learning

Robotron Datenbank-Software GmbH, Dresden.

Besuchen Sie unseren Vortrag:

21. Oktober 2020, 9:30 bis 10:00 Uhr

Dr. Deepa Kasinathan

Gruppenleiterin Computer Vision

# OPTISCHE QUALITÄTSPRÜFUNG UND PROZESSOPTIMIERUNG MITTELS DEEP LEARNING

Künstliche Intelligenz (KI), ein Kollektivbegriff, der die Fähigkeit von Maschinen beschreibt, eine menschenähnliche Kognition zu zeigen, hat sich als eine der umstrittensten, disruptivsten und transformativsten Geschäfts- und Technologietrends herauskristallisiert, mit dem Unternehmen heute konfrontiert sind. Länder, die diese technologischen Veränderungen vorantreiben und in industriellen Führungspositionen bleiben wollen, haben sich strategisch für die neue Art von cyber-physikalischer Infrastruktur positioniert, die aus der KI hervorgehen wird.<sup>[1,2,3]</sup> „Industrie 4.0“ ist eine nationale strategische Initiative der Bundesrepublik Deutschland, die darauf abzielt, eine wettbewerbsfähige Infrastruktur für die zukünftige Fertigungsindustrie sicherzustellen. Sie unterstützt die Integration von cyberphysikalischen Systemen und dem Internet der Dinge (Internet of Things, IoT) mit dem Ziel, die Produktivität, Effizienz und Flexibilität von Fertigungsprozessen und damit das Wirtschaftswachstum zu steigern.<sup>[4]</sup> Das „Industrie 4.0“ Rahmenprogramm hat sich zu einer europaweiten Zusammenarbeit entwickelt, um intelligente Automatisierung in großem Maßstab durchzuführen. Andere bemerkenswerte Beispiele, die in einem ähnlich Kontext begonnen wurden, sind die „Smart Manufacturing Leadership Coalition“ (SMLC) der Vereinigten Staaten, das „China Manufacturing 2025“-Programm und die Japanische „Society 5.0“.<sup>[5,6]</sup>

Die schnelle und genaue industrielle Inspektion zur Gewährleistung höchster Qualitätsstandards zu einem wettbewerbsfähigen Preis ist eine der größten Herausforderungen in der Fertigungsindustrie. Die Europäische Kommission schätzt, dass in einigen Industriezweigen 50 % der Produktion aufgrund von Mängeln aussortiert werden müssen und die Fehlerquote in komplexen Fertigungslinien bis zu 90 % erreichen kann.<sup>[7]</sup> Die manuelle Inspektion von Produkten, Teilen und Komponenten kann ein mühsamer und teurer Prozess sein. Erstens erfordert es eine beträchtliche Ausbildung für menschliche Experten, um solche Inspektionen durchzuführen. Zweitens können solche Inspektionen zu Engpässen bei den Produktions-/Zeitvorgaben für die Markteinführung führen. Drittens sind manuelle Inspektionen nicht so skalierbar wie Produkte. Die wissenschaftliche Disziplin namens **Computer Vision** (CV), die auf Bild- und Videoerkennung basiert, zeichnet sich als eine zielführende Alternative zur Lösung dieses Problems aus. Dank der möglichen Verfügbarkeit ausreichender Daten aus praktisch jedem Prozessschritt des Produktionsprozesses und der Entwicklung von Hardware, die leistungsfähig genug sind, um Berechnungen in Echtzeit durchzuführen, besteht für die Industrie ein großes Potenzial, die Fähigkeit von Maschinen zur Erkennung des eigenen Zustands durch intelligente Bildverarbeitung zu erhöhen. Eine CV-basierte automatisierte, optische Inspektion ermöglicht eine frühzeitige Fehlererkennung in den Herstellungsprozessen und trägt dazu bei, eine hohe Qualität der Produkte zu gewährleisten. Durch die genaue Erkennung der Fehlerklassen und die Sammlung historischer Produktionsstatistiken lassen sich Rückschlüsse auf den Produktionsprozess an sich ziehen. Damit wird eine Verbesserung der Fertigungslinien, eine Erhöhung des Automatisierungsgrades und eine Steigerung der Produktqualität ermöglicht.

Im industriellen Kontext bieten sich eine Reihe von Anwendungsfeldern für eine automatisierte Sichtprüfung mittels CV-Techniken. Die Maschine ist dabei in der Lage über einen langen Zeitraum, konstante Ergebnisse in der Beurteilung der Objekte zu liefern<sup>[8]</sup>. Anhand des Prüfziels lassen sich folgende Aufgabenklassen definieren<sup>[9]</sup>:

- Objekterkennung,
- Lageerkennung,
- Vollständigkeitsprüfung,
- Form- und Maßprüfung,
- Oberflächeninspektion.

Im Folgenden wird der Bereich CV näher vorgestellt und es wird skizziert, wie mittels KI die Leistungsfähigkeit des maschinellen Sehens auf ein neues Level gehoben werden kann.

## State of the art

Seit Beginn der 1970er Jahre versuchen Forscher im Feld Computer Vision die Fähigkeit des Menschen nachzuahmen, seine Umgebung mit Hilfe des visuellen Systems wahrzunehmen<sup>[10]</sup>. Aus einer Vielzahl von Daten und den erwarteten Ergebnissen wird ein Modell trainiert, welches Muster findet bzw. Regeln generiert, die auf neue, noch unbekannte Daten angewendet werden können. Über die Jahre haben sich in diesem Bereich verschiedene Techniken und Algorithmen entwickelt. Im Jahre 2012 wurde erstmals ein neuronales Netz entwickelt, welches an der Funktionsweise des menschlichen Gehirns angelehnt ist<sup>[11]</sup> und bei der ImageNet Large Scale Visual Recognition Challenge (ILSVRC) antrat<sup>[12]</sup>. Das sogenannte AlexNet [10] konnte diesen Wettbewerb mit deutlichem Vorsprung gegenüber Teams, welche von traditionellen CV-Verfahren Gebrauch machten, gewinnen. Durch diesen Meilenstein nahmen die Forschungsinteressen innerhalb der CV mittels Deep Learning (DL) und speziell unter dem Einsatz von sogenannten „Convolutional Neural Networks“ (CNN) stark zu. Auch die Entwicklungen im Hardware-Bereich haben diesen Trend positiv beeinflusst, da der Rechenaufwand sowohl beim Training als auch bei der Anwendung eines neuronalen Netzes enorm ist.

## Deep Learning im Bereich Computer Vision

Die CNNs bestehen im Kern aus drei verschiedenen Schichten (Convolutional Layer, Pooling Layer und Fully Connected Layer). Auf eine oder mehrere Convolutional Layer folgt ein Pooling Layer. Dieser Verbund wiederholt sich beliebig. Diese Schichten haben die Aufgabe die relevanten Ausschnitte, die Objekte beschreiben und ausmachen (sog. Features), aus dem Bild zu extrahieren. Die Convolutional Layer bestehen dabei aus Filter, die über das Bild laufen (z. B. zur Extraktion von Kanten) und deren Parameter trainierbar sind. Die Pooling-Schichten abstrahieren die Darstellungsform der Features, um deren Generalisierbarkeit zu erhöhen und sorgen für eine Verringerung der räumlichen Größe, was ebenfalls den Berechnungsaufwand reduziert. Die Fully Connected Layers sind anwendungsabhängig aufgebaut. Beispielsweise ist es bei der Bildklassifikation ihre Aufgabe, aus den Features der vorangegangenen Schichten eine Klassifizierung durchzuführen.

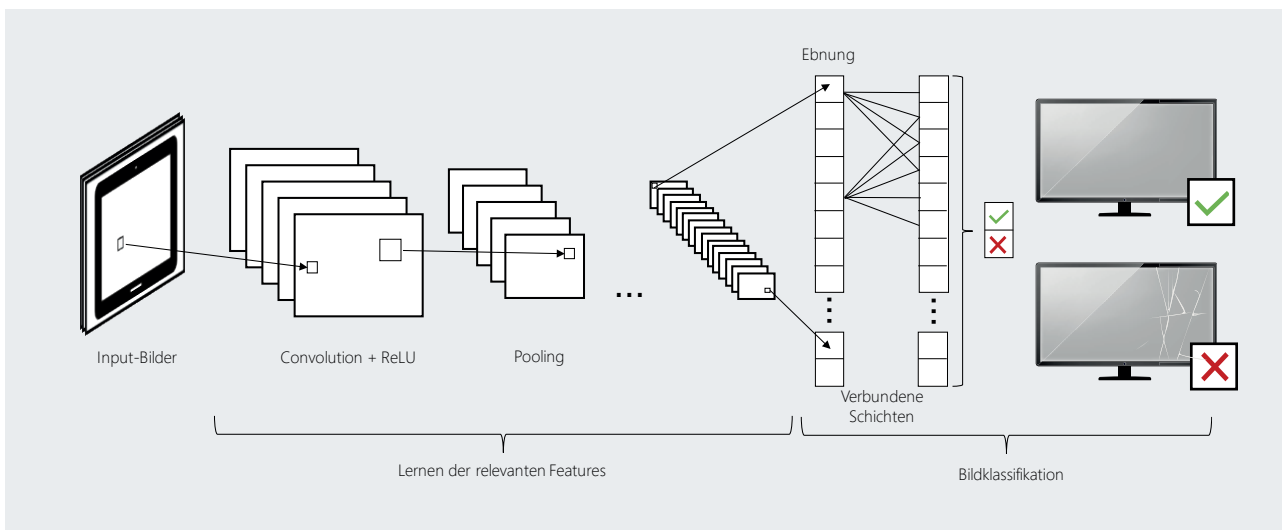


Abb. 1: Schematischer Aufbau eines CNNs zur Bildklassifikation (z. B. IO/NIO-Prüfung) (eigene Darstellung)

Das Beispiel zeigt den schematischen Aufbau eines CNNs zur Bildklassifikation. Es wird unterschieden, ob ein Monitor in Ordnung ist oder nicht den Qualitätsanforderungen entspricht (im Beispiel: Riss im Display). Das Bild wird an die erste Convolutional-Schicht übergeben. Innerhalb dieser Schicht laufen die Filter-Kernel über das Bild und bilden die Feature Maps (extrahierte Merkmale der Objektklassen). Im Anschluss folgt der Pooling Layer. Dieser Verbund wiederholt sich beliebig oft, bevor das Netz mit den Fully Connected Layern abschließt. Die Softmax-Funktion übernimmt am Ende die Klassifizierung dieses Zwei-Klassen-Problems. Für das dargestellte Bild sollte die ausgegebene Wahrscheinlichkeit der Softmax-Funktion für die Klasse „einwandfreier Monitor“ am höchsten sein.

## Traditionelle Verfahren vs. Deep-Learning-basierte Architekturen

Bei traditionellen CV-Verfahren stellt das Feature Engineering einen entscheidenden Teilschritt dar. Diese Features sind Ausschnitte aus dem Gesamtbild, die Objekte beschreiben und ausmachen. Dieser manuellen Extraktion erfordert ein hohes Maß an Domänenexpertise, denn der CV-Engineer muss entscheiden, welche Features für welche Objekte bzw. Klassen relevant sind. Der Vorteil der DL-basierten Verfahren liegt in ihrem End-to-End-Lernansatz. Man übergibt die Bilder und die gekennzeichneten Objekte im Bild und das neuronale Netz lernt in seinem Trainingsprozess die Muster und Features zu erkennen, welche die Objekte ausmachen. Dazu passt es sämtliche Parameter (bspw. in den Convolution-Schichten) an<sup>[13]</sup>. Das bereits erwähnte AlexNet hat rund 60 Millionen Parameter, die es während dieses Trainingsprozesses auf die gewünschten Bilddaten anzupassen gilt<sup>[14]</sup>. Die folgende Abbildung zeigt noch einmal den Unterscheid zwischen den Verfahrensweisen traditioneller Machine-Learning-Ansätze sowie Deep-Learning-Verfahren.

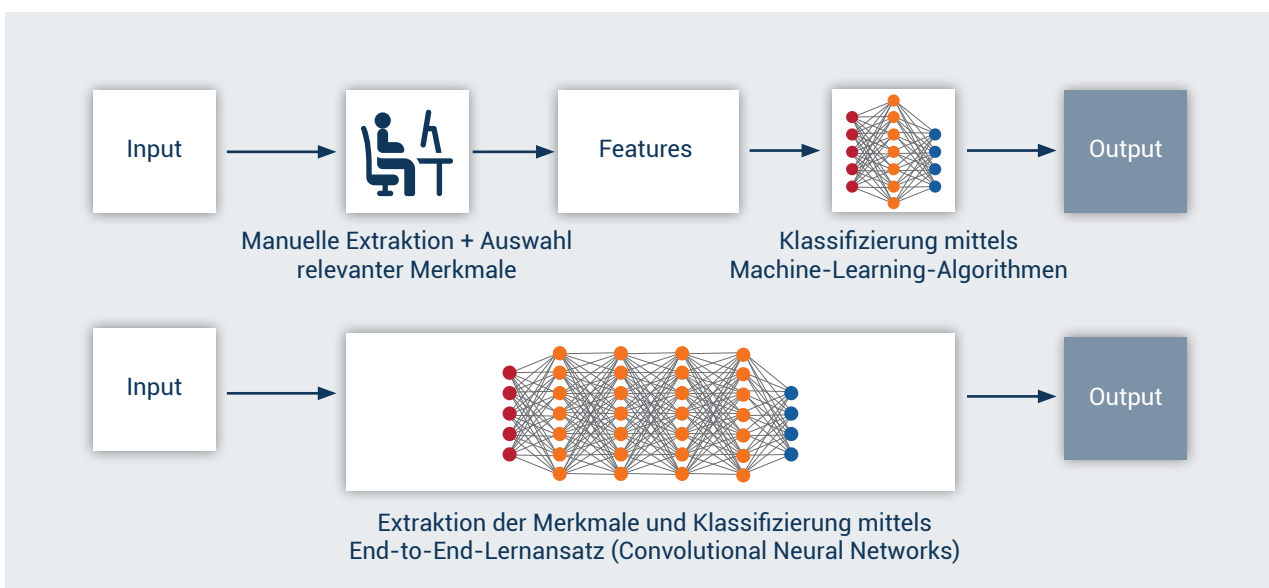


Abb. 2: Vergleich der Verfahrensweisen bei traditionellen Machine-Learning-Ansätzen und Deep-Learning-Verfahren (basiert auf<sup>[15]</sup>)

Nachfolgend sind sowohl die Vorteile als auch die Nachteile der CNN gegenüber den traditionellen ML-Ansätzen dargestellt.

### Vorteile

- Durch die Bereitstellung verschiedenster Bilder im Trainingsprozess arbeiten neuronale Netze robuster und flexibler gegenüber Veränderungen während der Bildaufnahme (bspw. andere Lichteinflüsse, andere Produktionslinien), während CV-Experten oft ein Finetuning vornehmen müssen oder gar bisher angewendete Verfahren komplett umgestellt werden müssen.
- Weniger Domänenexpertise notwendig. Das neuronale Netz extrahiert relevante Merkmale von allein im End-to-End-Lernprozess.
- Sie sind mittlerweile State of the Art in sämtlichen Benchmarks im Bereich Computer Vision.

### Nachteile

- Der Trainingsprozess erfordert eine hohe Rechenleistung, was heutzutage allerdings nur noch geringe Probleme darstellt.
- Das Fachwissen von CV- und Domänenexperten verschiebt sich hin zu Kenntnissen im Bereich der künstlichen Intelligenz, wobei dieser Prozess zunehmend automatisiert wird und nur noch die Bilder von entsprechenden Prozessexperten für das Training annotiert werden müssen.
- Für manche Anwendungsfälle sind DL-basierte Architekturen evtl. überdimensioniert und traditionelle Verfahren könnten diese effizienter lösen.

## Schritt für Schritt zur Anwendung

Der typische Prozess bei der Erstellung eines Modells zur Verarbeitung von Bilddaten ist im Folgenden abgebildet. Zunächst wird die zu bearbeitende Aufgabe abgegrenzt. Dazu zählen bspw. die Vollständigkeitsprüfung oder wie im Beispiel die Kontrolle auf prozessbedingte Oberflächenfehler (Erkennung von Rissen). Die Abbildung zeigt verschiedenen Schritte eines End-to-End Computer-Vision-Projektes, beginnend mit der Datenaufbereitung bis hin zum Deployment des finalen Modells in die Inferenz-Umgebung.

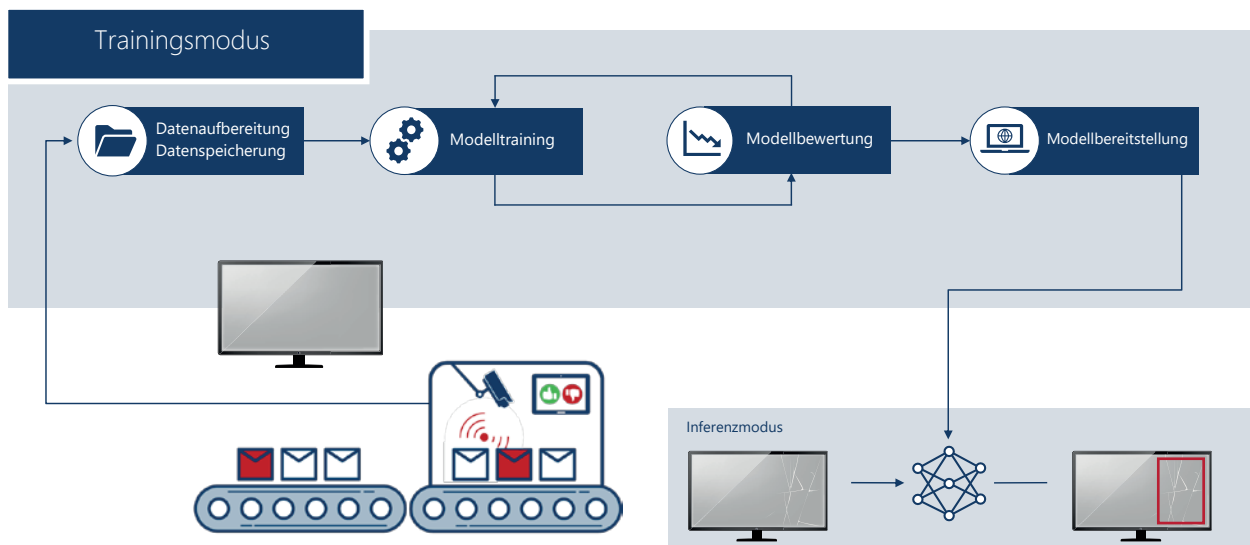


Abb. 3: Die verschiedenen Schritte eines Computer-Vision-Projektes (Quelle: Robotron Datenbank-Software GmbH)



## Bildaufnahme und Datenvorbereitung

Innerhalb des Einsatzgebietes werden die entsprechenden Bilder von den Prüfteilen aufgenommen. Dabei ist es wichtig, sowohl eine große Auswahl der Fertigungsrahmenbedingungen (z. B. Lichtszenarien in den Schichten), als auch eine ausreichende Anzahl an Bildern der jeweiligen Klassen zu integrieren. Wie anfangs beschrieben, werden dem Modell in der Trainingsphase aber nicht nur die Bilder, sondern auch die entsprechenden Annotationen übergeben. Dieser Schritt ist entscheidend, da die Modellgüte von der bereitgestellten „Ground Truth“ abhängig ist. Die Ground Truth beschreibt dabei den vom Domänenexperten annotierten Bilddatensatz. Das Modell wird anhand der Übereinstimmung mit dieser optimiert. Um die Annotationen zu erzeugen, werden sogenannte Labeltools genutzt. In der Abbildung ist das von der Robotron Datenbank-Software GmbH entwickelte Labeltool (Realtime Computer Vision, **RCV**) dargestellt. In dem erkennbaren Ausschnitt werden die im Bild befindlichen Objekte (im dargestellten Fallbeispiel handelt es sich um einen Riss) markiert, um im Anschluss durch das neuronale Netz diese zu detektieren und über die Anzahl eine entsprechende Vollständigkeitsprüfung durchzuführen. Insbesondere das manuelle Labeling großer Bilddatensätze ist eine monotone und zeitaufwändige Aufgabe. Abhilfe können hier der Einsatz moderner und flexibler Labeling-Tools schaffen. So können etwa Funktionen zur Unterstützung insbesondere von lokaler Annotation die Labelingdauer auf ein Fünftel verkürzen, was die Wirtschaftlichkeit insgesamt deutlich verbessert. Zum Beispiel ist es möglich, dass vortrainierte Modelle bereits Labels vorschlagen, die der Bearbeiter dann nur noch bestätigen und ggf. korrigieren muss („semiautomated labeling“) oder, dass Bildverarbeitungsmethoden bei der automatischen Auswahl von relevanten Regionen unterstützen. Die folgende Abbildung zeigt einen Bildausschnitt aus dem Robotron Labeltool.

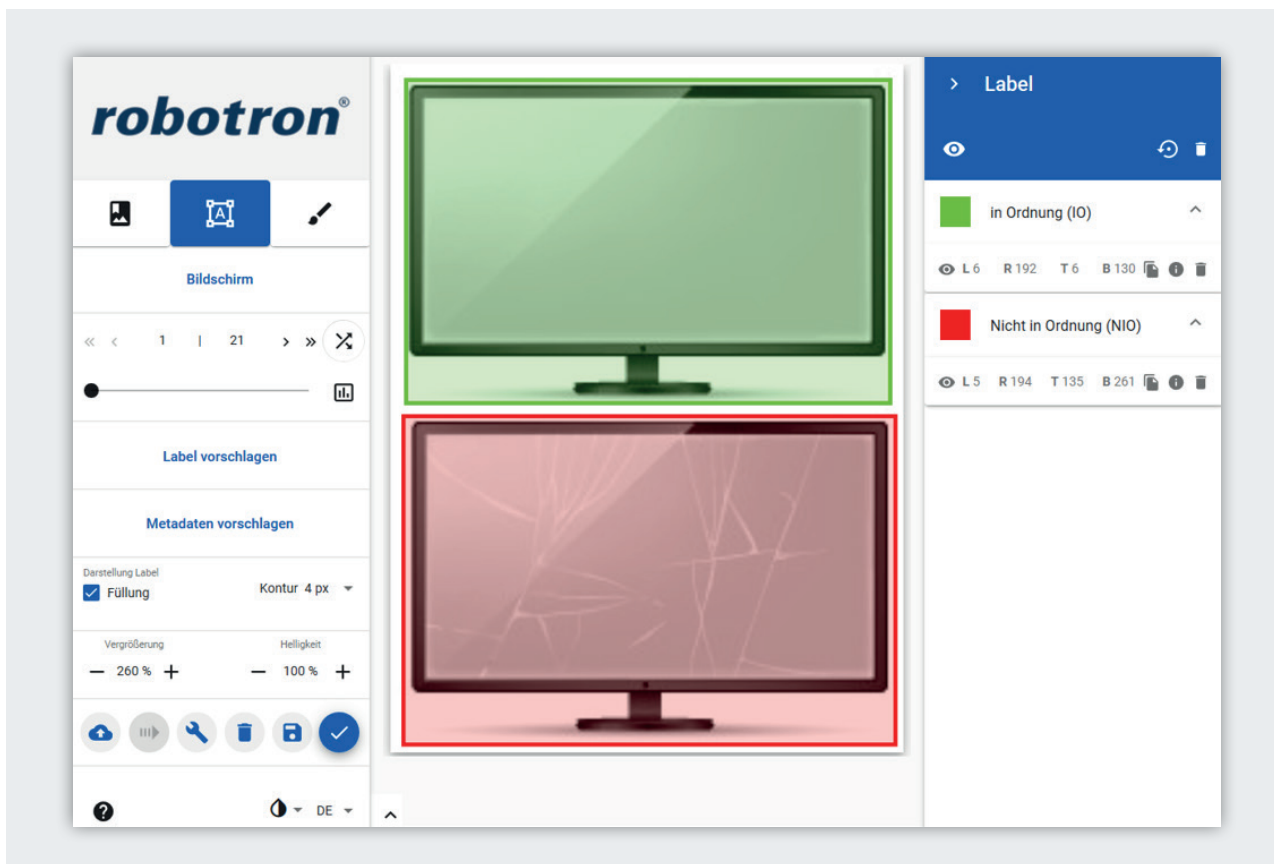


Abb. 4: Beispiel für ein Labeltool zum Annotieren von Trainingsbildern  
(Quelle: Realtime Computer Vision – Labeltool von Robotron Datenbank-Software GmbH)

## Modelltraining und Bewertung

Nachdem eine ausreichende Zahl an Bildern annotiert wurde, kann mit dem Modelltraining begonnen werden. Das Training eines Modells erfordert großen Rechenaufwand, sodass der Vorgang in der Praxis überwiegend in der Cloud stattfindet. Die Auslagerung dieses zwar initial nötigen, aber i. d. R. nur einmal nötigen ressourcenhungrigen Vorgangs in die Cloud, ist meist wirtschaftlicher als die Anschaffung und der Betrieb leistungsstarker Server für diesen Zweck. Es gibt verschiedene Frameworks, die diverse Modellarchitekturen unterstützen. Es gibt eine Reihe von Deep-Learning-Bibliotheken mit denen sich neuronale Netze erstellen lassen. Außerdem gibt es verschiedene Architekturen, die es abhängig von der zugrundeliegenden Aufgabenstellung zu wählen gilt. Während des Trainings werden sogenannte Evaluationsmetriken erhoben, mit denen sich die Güte eines Modells bewerten lässt (Validität). Diese sind ausschlaggebend dafür, wann der Trainingsprozess beendet wird. Bevor das Modell zum Einsatz kommt, sollte es ausführlichen Tests unterworfen werden, wie vor allem einer Analyse der Generalisierbarkeit („Robustness“) sowie einem Test von Grenzfällen.

## Modellbereitstellung/Inferenzmodus

Sofern ein Modell die gewünschte Güte erreicht hat, kann es bereitgestellt werden. Je nach Komplexität des Modells ist eine Implementierung bereits auf einfacher Hardware bzw. eine Integration in bestehende Steuerungssysteme möglich. Nachdem die Kamera angeschlossen und geeignet positioniert wurde, kann mit der Inferenz begonnen werden. Als Inferenz bezeichnet man den Prozess der Klassifizierung bzw. Erkennung und Bewertung des Kamerabildes durch das Modell. Bei einer globalen Klassifikation ist das Ergebnis der Inferenz eine oder mehrere Klassen zum Bild, während bei einer lokalen Objekterkennung auch jeweils die genaue Position und Form der Objekte erkannt wird (siehe Abbildung 5).

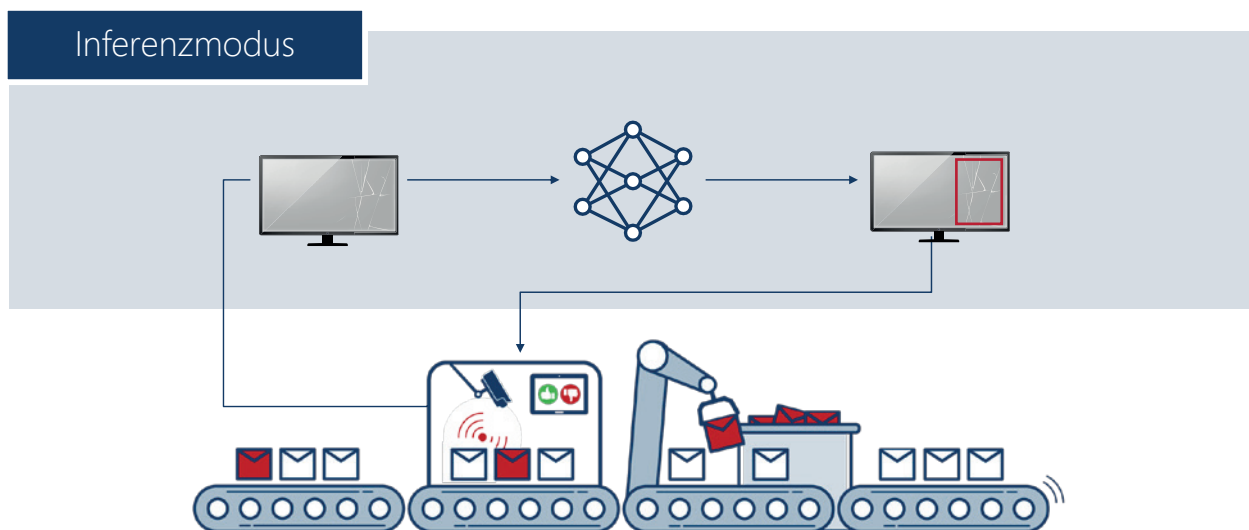


Abb. 5: Schematische Darstellung eines Computer Vision Einsatzes in der Industrie.  
(Quelle: Robotron Datenbank-Software GmbH)

Nachdem die Bereitstellung des Modells abgeschlossen ist, können die Ergebnisse der Inferenz überwacht werden (siehe Abbildung 6). Durch die kontinuierliche visuelle Prüfung ist es darüber hinaus nun ebenfalls möglich, bei Erkennung verdächtiger Muster automatisch und frühzeitig Warnungen auszugeben.

Die Ausschussrate kann so erheblich reduziert und Folgefehlern vorgebeugt werden. Ist das Modell einmal trainiert und vor Ort eingerichtet, wird es im konkreten Anwendungsfall kontinuierlich Daten liefern, ohne dass ein Mitarbeiter benötigt wird. Zudem lässt sich das Modell jederzeit weiterentwickeln und z.B. an eine neue bzw. veränderte Prozessumgebung anpassen, indem es mit neuen annotierten Bildern aktualisiert bzw. nachtrainiert wird.

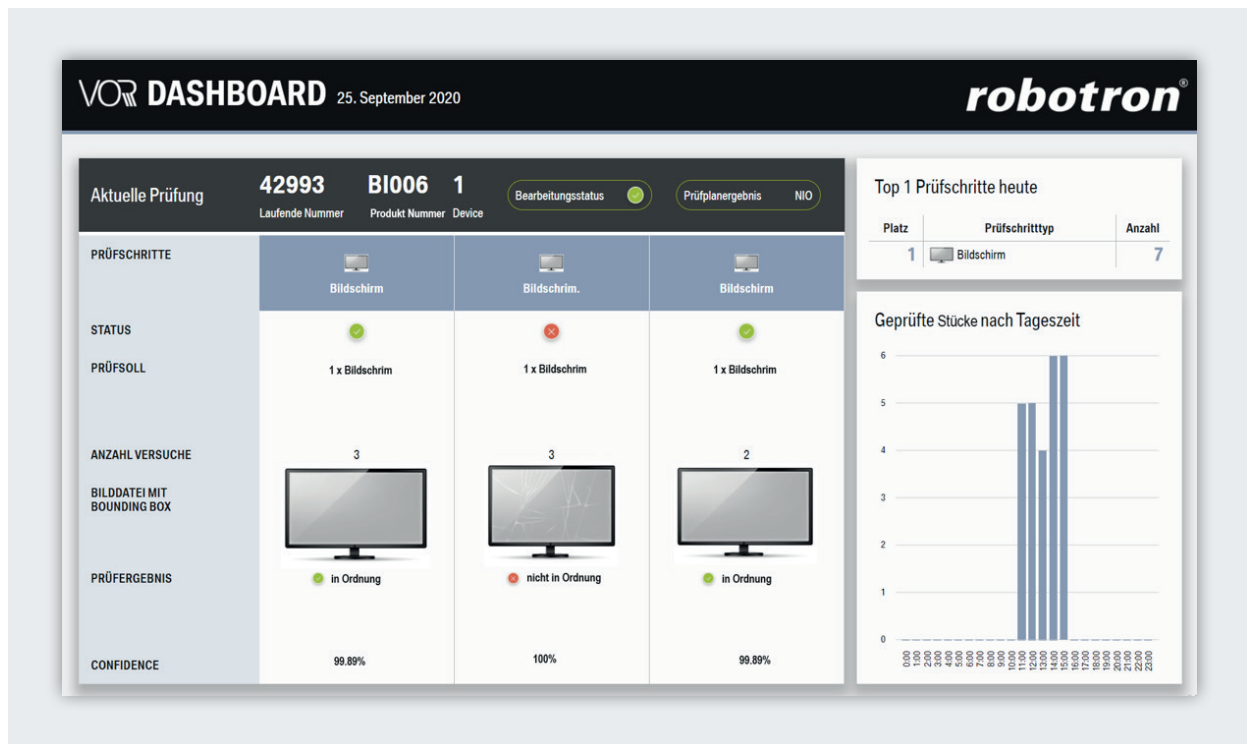


Abb. 6: Beispiel für eine Inferenz-Visualisierungsoberfläche.

(Quelle: Realtime Computer Vision – VOR Dashboard von Robotron Datenbank-Software GmbH)

## Fazit

Im Hinblick darauf, dass Industrie 4.0 eine größere Abhängigkeit von maschinellem Lernen vorausieht, spielt die Computer-Vision-Technologie eine wesentliche Rolle bei der Erfassung der physischen Welt und ihrer Umwandlung in vernetzte, digitale Daten, wodurch eine breite Palette von Produkten ermöglicht wird, die intelligenter, reaktionsschneller sowie robuster gegen äußere Einflüsse als bisher und damit wertvoller für die Anwender sind. Vor allem die Informationen in den Daten der Bilder können nahezu in Echtzeit auf diverse komplexe Fragestellungen hin analysiert und verarbeitet werden, um die Produktivität zu steigern, Qualitätsstandards aufrechtzuerhalten und Verschwendung zu reduzieren, so dass Unternehmen in einem Kreislauf der kontinuierlichen Verbesserung ihre betriebliche Effizienz steigern können. Darüber hinaus stellt diese Technologie die Verbindung zwischen Maschinenautomatisierung, Informationsautomatisierung und Wissensautomatisierung her. Letztendlich können durch die Umwandlung von Daten in Erkenntnisse sowie den daraus resultierenden faktenbasierenden Maßnahmen die Produktivität und Effizienz in der Fertigung in naher Zukunft nachhaltig gesteigert werden. In der heutigen Zeit steht eine weltweite Industriallianz zur Verfügung, um die industrielle Fertigung dabei zu unterstützen, die Computer-Vision-Technologie optimal in ihre Hardware- und Software-Designs zu implementieren.



## Über Robotron Datenbank-Software GmbH



Die Robotron Datenbank-Software GmbH begleitet Industrieunternehmen bei der Umsetzung der Digitalen Transformation – von der ersten Idee über die Beratung, Analyse und Entwicklung innovativer Lösungen bis zur agilen Umsetzung.

Auf dem Weg zum IIoT und der Connected Factory verstehen wir Maschinen- und Prozessdaten als wertvolle Basis für die Prozessführung und -optimierung. Mittels aktueller Edge- und Cloud-Technologien, Künstlicher Intelligenz sowie analytischer und methodischer Expertise können wir Daten nicht nur sammeln, überwachen und visualisieren, sondern auch hinsichtlich konkreter Fragestellungen auswerten und Handlungen ableiten. Auf diese Weise generieren wir entscheidende Mehrwerte für die Produktionsprozesse unserer Kunden.

Mit der **RCV**-Plattform bieten wir eine modulare Lösung für sämtliche Computer Vision Anwendungsfälle wie Bildklassifikation, Objekterkennung sowie Bildsegmentierung unter Verwendung von Künstlicher Intelligenz, neuronalen Netzen und Maschinellen Lernen.

Haben Sie weitere Fragen?

**Dann besuchen Sie unsere Website:**

<https://www.robotron.de/produkte/rcv/>

## Quellenangaben:

- [1]. Scaling AI in Manufacturing Operations: A Practitioners' Perspective, Capgemini Research Institute (2019) <https://www.capgemini.com/gb-en/wp-content/uploads/sites/3/2019/12/Report-%E2%80%93-AI-in-Mfg-Ops.pdf>
- [2]. Driving impact at scale from automation and AI, Digital/McKinsey (2019) <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Driving%20Impact%20at%20Scale%20from%20Automation%20and%20AI/Driving-impact-at-scale-from-automation-and-AI.ashx>
- [3]. Potenziale der Künstlichen Intelligenz im Produzierenden Gewerbe in Deutschland, iit-Institut für Innovation und Technik in der VDI / VDE Innovation + Technik GmbH (2018) [https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/potenziale-kuenstlichen-intelligenz-im-produzierenden-gewerbe-in-deutschland.pdf?\\_\\_blob=publicationFile&v=8](https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/potenziale-kuenstlichen-intelligenz-im-produzierenden-gewerbe-in-deutschland.pdf?__blob=publicationFile&v=8)
- [4]. <https://www.plattform-i40.de/PI40/Navigation/DE/Home/home.html>
- [5]. Li, L. China's manufacturing locus in 2025: With a comparison of 'Made-in-China 2025' and 'Industry 4.0'. Technological Forecasting and Social Change, vol. 135, pages 66-74, (2018)
- [6]. Shiroishi, Y.; Uchiyama, K.; Suzuki, N. Society 5.0: For Human Security and Well-Being. Computer 51, S. 91–95, (2018)
- [7]. <https://ec.europa.eu/programmes/horizon2020/en/news/intelligent-approach-zero-defect-manufacturing>
- [8]. Demant, C., Streicher-Abel, B. und Springhoff, A. (2011). Industrielle Bildverarbeitung. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [9]. Steger, C., Ulrich, M. und Wiedemann, C. (2008). Machine vision algorithms and applications. Publication Title: Textbook. Weinheim: WILEY-VCH.
- [10]. Szeliski, R. (2011). Computer vision algorithms and applications. London , Heidelberg[u.a.]: Springer.
- [11]. Aggarwal, C. C. (2018). Neural Networks and Deep Learning: A Textbook. Springer International Publishing.
- [12]. ILSVRC ist ein Wettbewerb, indem Algorithmen zur Objekterkennung und Bildklassifikation evaluiert werden. (<http://www.image-net.org/challenges/LSVRC/>)
- [13]. Walsh, J., O' Mahony, N., Campbell, S. et al. (2019). "Deep Learning vs. Traditional Computer Vision".
- [14]. Krizhevsky, A., Sutskever, I. und Hinton, G. E. "ImageNet Classification with Deep Convolutional Neural Networks". In: Advances in Neural Information Processing Systems 25. Hrsg. von F. Pereira, C. J. C. Burges, L. Bottou et al. Curran Associates, Inc., S. 1097–1105 (2012)
- [15]. Wang, J., Ma, Y., Zhang, L. et al. "Deep Learning for Smart Manufacturing: Methods and Applications". In: Journal of Manufacturing Systems 48, S. 144–156 (2018)