

Computer Vision und Deep Learning

Know-how Komplexe Entscheidungsprozesse erfordern oft erweiterte Techniken, welche über herkömmliche Methoden des maschinellen Sehens hinaus gehen.

Von Jan Riedo und Marcel Messerli

In der Industrie trägt der Einsatz von Machine-Vision-Verfahren (maschinell Sehen, Computer Vision) seit Jahrzehnten zur Effizienzsteigerung in der Qualitätskontrolle bei. Allerdings beschränken sich die Automatisierungsmethoden in der Regel auf einfache visuelle Inspektionsaufgaben wie Grössenausmessungen, Lage-Erkennung oder die Erkennung von einfachen Identifikationsmerkmalen. Bei komplexen Entscheidungsprozessen in der Qualitätssicherung und anspruchsvollen Klassifizierungen stossen herkömmliche Machine-Vision-Methoden aber an ihre Grenzen. Eine Kontrolle der Fabrikate ist unter diesen Umständen lediglich mit durch den Menschen vorgenommenen Stichproben und den entsprechenden

Kostenfolgen möglich. Deep Learning erlaubt nun aber mittlerweile auch anspruchsvollere Kontrollen von Fabrikaten, etwa bei zufälliger Anordnung, zufälligen Formen oder sich ändernden Beurteilungskriterien. Mit Deep Learning sind indes nicht nur visuell ambitionierte Kontrollen möglich. Es erlaubt auch, flexibel auf kurzfristig auftretende, neue Anforderungen und Marktsituationen zu reagieren und so die Wettbewerbsfähigkeit von Unternehmen zu steigern. Im Folgenden soll gezeigt werden, für welche Aufgaben bei Entscheidungsprozessen ein Vision-System mit Deep Learning eingesetzt werden kann, wie es aufgebaut ist und welches die Bedingungen für einen erfolgreichen Einsatz sind.

Automation von komplexen Aufgaben

Seit der Industrialisierung sind Maschinen in der Produktion von Gütern nicht mehr wegzudenken. Dabei werden diese heutzutage nicht nur als Unterstützung der Fabrikarbeitenden eingesetzt, sondern können auch grosse Teile des Produktionsprozesses vollständig und autonom übernehmen. In diesen Prozessen ist eine Kontrolle der Qualität der Fabrikate sehr wichtig. In vielen Fällen ist eine Qualitätssicherung durch visuelle Begutachtung durchzuführen. Obwohl für verschiedenste Industriezweige anwendbar, soll als Anschauungs-Beispiel die Biskuit-Herstellung aus der Lebensmittelindustrie dienen. Dort muss sichergestellt werden, dass die Kekse nicht schadhafte sind oder etwa der Schokoladenüberzug keine Fehler aufweist. Die visuelle Inspektion jedes einzelnen Fabrikates ist bei einem Prozess mit hohem Automatisierungsgrad und entsprechend hohem

Durchsatz nur automatisiert möglich. Doch welche Inspektionsaufgaben lassen sich überhaupt automatisieren?

Bei klassischem Machine Vision handelt es sich um Aufgaben, bei welchen die gewünschte Information aus dem Bild Pixel für Pixel extrahiert und mit dem Sollwert verglichen wird, ohne dass die Informationsextraktion über Zeit gelernt wurde. Solche Aufgaben sind zum Beispiel Ausmessungen (Grösse des Kekses), Vollständigkeitsprüfungen (Wie viele Stück sind in die Packung abgefüllt worden?), Erkennung von Identifikationsmerkmalen (z.B. hell oder dunkel) oder die Lage-Erkennung (z.B. im Hoch- oder Querformat auf dem Förderband). Über klassische Machine-Vision-Aufgaben hinaus gibt es aber auch Probleme, welche erst durch den Einsatz von Deep-Learning-Methoden, die in modernen KI-Anwendungen integraler Bestandteil sind, effizient gelöst werden können. Im Gegensatz zum klassischen maschinellen Lernen bietet Deep Learning ein flexibles Modell an, mit dem sich Informationen aus dem Bildmaterial ableiten lassen, welche nicht direkt aus den Pixel-für-Pixel-Informationen ersichtlich sind. Dies ist überall dort der Fall, wo die Ansicht auf die Fabrikate nicht konstant gehalten werden kann (zufällige Anordnung) oder die einzelnen Fabrikate sich stark voneinander unterscheiden.

Mit der Wahl eines Deep-Learning-Ansatzes steht infolgedessen das Training des Modells im Vordergrund. Das bedeutet, dass eine Datenbasis aufgebaut wird und neben den Gütern selbst auch Daten produziert werden. Im Folgenden soll aufgezeigt werden, für welche Aufgaben ein Vision-System mit Deep Learning verwendet werden kann, wie es aufge-

DIE AUTOREN

Jan Riedo ist Data Scientist und Data & Analytics Consultant bei IT-Logix. Er hält einen Master in Bio-medical Engineering mit Spezialisierung im Bereich intelligente Bildverarbeitung mit Deep Learning. Er konnte in diversen Branchen Data-Science- und Data-Warehousing-Lösungen umsetzen.



Marcel Messerli ist Senior Data Scientist und Data & Analytics Consultant bei IT-Logix. Er hält einen Master in Mathematik und hat mehrjährige Praxiserfahrung in Data Science, Data Engineering und Software-Entwicklung.



baut ist und welche Voraussetzungen gelten müssen, damit es erfolgreich eingesetzt werden kann.

Anspruchsvolle Vision Tasks

Industrielle Machine-Vision-Anwendungen gibt es schon seit mehr als 20 Jahren. Auf dem Markt tummeln sich daher auch sehr viele Anbieter von Soft- und Hardware-Lösungen für Inspektionsaufgaben. Diese haben sich für die klassischen Vision-Tasks sehr gut bewährt. Eine wichtige Voraussetzung für ein reibungsloses Funktionieren dieser althergebrachten An-

sätze ist indes die Gleichmässigkeit der Rahmenbedingungen. Bei einer Klassifikation beispielsweise müssen die Fabrikate immer exakt in der gleichen Position aufgezeichnet werden. Oder es muss etwa bei der Fehlererkennung ein Algorithmus definiert werden, der alle Fehler aus dem Bild erkennen kann. Dieser Algorithmus basiert dann auf einer Wenn-Dann-Sonst-Logik, für welche alle möglichen Fehler schon bekannt und definiert sein müssen.

Es gibt allerdings Vision-Projekte, bei denen die gängigen Software-Lösungen nicht zum Ziel führen oder ein sehr

grosser Aufwand nötig wäre, weil die notwendige Unveränderlichkeit der Rahmenbedingungen nicht gegeben ist. Für solche und ähnliche Probleme kann eine eigens dafür entwickelte Lösung, basierend auf Deep Learning, eine gute Alternative bieten. Die Grundlage von Deep Learning sind neuronale Netzwerke, welche sich von der Architektur her am menschlichen Gehirn orientieren. Anders als bei einem klassischen Machine-Vision-Problem werden für das Trainieren des Modells Daten benötigt. Um bei unserem Beispiel zu bleiben, werden Fo-

BAUSTEINE DER BILDVERARBEITUNG MIT DEEP LEARNING

1) KAMERA

Industriekameras gibt es in verschiedensten Auflösungen und Schutzklassen. Für klassische Machine-Vision-Anwendungen wird üblicherweise eine monochrome Kamera verwendet, da Farben oftmals keine Rolle spielen. Bei Klassifizierungsaufgaben mit Deep Learning liefert die Farbe jedoch wichtige zusätzliche Informationen. Auf eine hohe Auflösung kann meist verzichtet werden, da die verwendeten statistischen Modelle mit kleineren Bildern einfacher zu trainieren sind. Die meisten Hersteller implementieren den GenICam-Standard (Generic Interface for Camera). Damit lässt sich die Kamera unabhängig von Hersteller, Firmware und Schnittstellentechnologie einheitlich steuern. Die gängigsten Hardware-Schnittstellen sind USB3 Vision und GigE Vision. Letztere wird bei längerem Kabel zwischen Kamera und Host bevorzugt.

2) BELEUCHTUNG

Am optimalsten ist eine kontrollierte Beleuchtung in einer von anderen Lichtquellen abgeschirmten Box. Falls dies nicht praktikabel ist, empfiehlt sich eine zusätzliche, verstärkte Beleuchtung. Die Ausleuchtung der Objekte ist bei klassischen Machine-Vision-Anwendungen viel wichtiger. Jedoch ist auch bei Deep-Learning-Ansätzen zu beachten, dass die Ausleuchtung einigermaßen konstant bleibt, da ansonsten bei Schattenbildung das Modell das Bild falsch interpretieren kann. Sind jedoch in den Daten zum Training die verschiedenen Ausprägungen der Ausleuchtung genügend gut repräsentiert, kann das Modell mit diesen umgehen.

3) TRIGGER

Je nach Problem werden die Bilder der Kamera nicht kontinuierlich benötigt, sondern lediglich, wenn sich ein Objekt darin befindet. Dies lässt sich mit Hilfe von Software detektieren. Vielfach ist es jedoch einfacher, wenn die Aufnahme durch einen Trigger, wie zum Beispiel einer Lichtschranke, ausgelöst wird. Häufig bieten die Kameras schon integrierte Schnittstellen, mit denen das Trigger-Signal abgegriffen werden kann.

4) BILDVERARBEITUNGS-SERVER / EDGE-SYSTEM

Wenn das Bild über die physische Schnittstelle zum Bildverarbeitungs-Server gelangt ist, ist der Treiber der Kamera für die Bereitstellung des Bildes verantwortlich. Die Funktionen zur Bild-Akquise können über eine API (GenTL Producer) in die eigene Applikation integriert werden. Die Verarbeitung der Bilder wird auf dem Server durchgeführt, wobei je nach Kamera bereits auf ihr selbst eine Vorverarbeitung erfolgen kann. Sind die Fotos im gewünschten Format vorhanden, wird das Modell auf diese angewendet und Bilder und Re-

sultate werden abgespeichert. Die Fotos dienen später als Trainingsdatensatz, um das Modell zu verfeinern.

5) MODELL-TRAINING

Dem statistischen Modell und dessen Training kommt eine besondere Rolle zu. Anders als bei der klassischen Machine Vision werden hier Parameter nicht manuell optimiert, sondern Millionen von Variablen innerhalb des neuronalen Netzwerks automatisch auf den bestmöglichen Wert justiert. Dem Netzwerk werden dazu jeweils die Bilder, etwa von Keksen mit Labeln («gut» / «schlecht») gezeigt. Das Netz lernt nun aus den Bildern eine Repräsentation für ein gutes und ein schlechtes Guetzli und berechnet aus neuen Bildern direkt die entsprechenden Label. Dieser zum Teil wenig stabile Prozess kann mitunter mehrere Tage dauern und birgt ganz eigene Herausforderungen. Entsprechend kann gezieltes Eingreifen bei der Bildvorverarbeitung, den Trainingsparametern und der Netzwerkarchitektur nötig sein. Es empfiehlt sich, das Modell nicht auf demselben Server zu trainieren, auf dem auch die operative Bildverarbeitung und die Anwendung eines Modells gerechnet wird. Gewisse Cloud-Dienste wie etwa Azure Machine Learning von Microsoft unterstützen die Versionierung und den Trainings- und Deployment-Prozess der neu gewonnenen Modelle.

6) GUI

Es ist sinnvoll, auf einem Bildschirm darzustellen, was das System gerade «sieht». Wird ein intelligentes Vision-System mit Supervised Deep Learning trainiert, werden zudem Bilder in gelabelter Form benötigt. Das heisst, dass zu jedem Bild eines Kekses auch die Zuordnung zur Qualität wie «gut» oder «schlecht» vorliegt, um das Modell zu trainieren. Für die Transparenz und die Akzeptanz des Systems ist ein GUI vor allem bei der Einführung des Vision-Systems sehr wichtig. Meist müssen bestehende Prozesse Schritt für Schritt transformiert werden. Wo neu Daten gesammelt werden, ergibt sich oft zusätzlich die Möglichkeit, Prozessinformationen zu extrahieren. Diese können in Business-Intelligence-Anwendungen in einem Dashboard dargestellt werden.

7) MENSCHLICHE KONTROLLE

Ein Vision-System, welches die Qualitätskontrolle übernehmen soll, muss selbst auch kontrolliert werden. Der Vorteil des Systems ist es, dass es Bilder für die menschliche Inspektion definieren kann. Zum einen sind das zufällig ausgewählte Fälle, zum anderen aber auch solche, in denen das Modell wie zum Beispiel bei einem Guetzli mit korrekter Schokoglasur aber abgebrochenen Ecken eine gewisse Unsicherheit zeigt. Schliesslich soll gewährleistet sein, dass sich das System ständig verbessern kann.