



Whitepaper

IT-LOGIX
BUSINESS INTELLIGENCE

IT-LOGIX AG

Schwarzenburgstrasse 11
3007 Bern

T +41 (0)848 848 058

F +41 (0)848 848 059

www.it-logix.ch



QUALITÄTSSICHERUNG IN DER INDUSTRIE MIT MACHINE VISION UND DEEP LEARNING

Wie Deep Learning und intelligente Inspektionsverfahren anspruchsvolle Klassifizierungen und Qualitätssicherung ermöglichen, Effizienz und Qualität in Produktionsprozessen steigern und sich dadurch signifikante Kosteneinsparungen erzielen lassen.

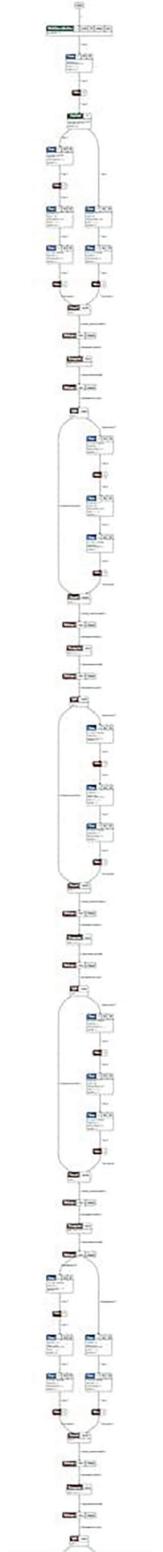




INHALT

	Seite
➤ Management Summary	3
➤ Motivation: Automation von komplexen Aufgaben	3
➤ Anspruchsvolle Vision Tasks in der Lebensmittelindustrie	3
➤ Prozess der Bildverarbeitung	4
➤ Gelabelte Daten	5
➤ Zusammenfassung	6
➤ Die Autoren	6

Ausschnitt aus dem Schema eines neuronalen Netzwerks mit sich wiederholenden Abschnitten, sogenannten Residualen Blöcken. Netzwerke dieser Art (sog. Residual Neural Networks) erzielen aktuell die besten Werte in der Bildklassifizierung.





Management Summary

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

Deep Learning, ein Teilbereich der Künstlichen Intelligenz (KI), erlaubt mittlerweile auch anspruchsvollere Kontrollen von Fabrikaten, etwa bei zufälliger Anordnung, zufälligen Formen oder sich ändernden Beurteilungskriterien. Mit Deep Learning sind 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.

Die vorliegende Publikation gibt unter Einbezug eines praxisnahen Beispiels (Qualitätskontrolle von Biscuits) interessierten CEOs, COOs, CFOs und Qualitätsverantwortlichen der Lebensmittel- und anderen Industriezweigen eine Einführung in die Grundlagen und Methoden von Deep Learning und intelligenter Bildverarbeitung bei anspruchsvollen Vision-Tasks. Es 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.

Motivation: 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, wird als Anschauungs-Beispiel die Guetzli (Biskuit-)Herstellung aus der Lebensmittelindustrie dienen. Dort muss sichergestellt werden, dass die Guetzli nicht schadhaft sind oder z.B. 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?



Machine Vision (maschinelles Sehen) bezeichnet industrielle Anwendungen, welche mit Hilfe von visuellen Systemen automatisierte Prozesse lenken. Hauptsächlich werden die Anwendungen in der Qualitätssicherung (QS) und in der Automatisierungstechnik eingesetzt.

Bei der klassischen Machine Vision (maschinelles Sehen) 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 (Wie gross ist das Guetzli?), Vollständigkeitsprüfungen (Wie viele Guetzli sind in die Packung abgefüllt worden?), Erkennung von Identifikationsmerkmalen (War es ein helles oder dunkles Guetzli?) oder die Lageerkennung (Ist das Guetzli im Hoch- oder im Querformat auf dem Förderband?). Über die klassischen 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 flexibleres Modell an, mit dem sich Informationen aus dem Bildmaterial ableiten lassen, welche nicht direkt aus der «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 eine zufällige Form aufweisen und sich also stark voneinander unterscheiden. Im Falle der visuell begutachteten Guetzli können dies komplexe Qualitätsmerkmale wie z.B. der Glanz der Schokolade sein.



Deep Learning ist eine Methode des machine learnings (maschinellen Lernens), welche meist in künstlich intelligenten Systemen Verwendung findet. Hierzu werden an neuronale Netzwerke, wie sie im Gehirn vorkommen, angelehnte Modelle benutzt. Wegen der Mächtigkeit der Methoden menschenähnliche Ergebnisse zu erzielen, wird Deep Learning in der Bilderkennung häufig eingesetzt.

Mit der Wahl eines Deep-Learning-Ansatzes steht infolgedessen das Training des Modells im Vordergrund. Das bedeutet, dass eine Datenbasis aufgebaut wird und neben der eigentlichen Produktion der Güter (hier: Guetzli) auch Daten «produziert» werden. In diesem Whitepaper soll aufgezeigt werden, für welche Aufgaben ein Vision-System mit Deep Learning verwendet werden kann, wie es aufgebaut ist und was für Voraussetzungen gelten müssen, damit es erfolgreich eingesetzt werden kann.

Anspruchsvolle Vision Tasks in der Lebensmittelindustrie

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 Ansätze 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 erwähnte notwendige Unveränderlichkeit der Rahmenbedingungen nicht gegeben ist. Bei einem Guetzli kann zum Beispiel der Schokoguss jedes Mal anders aussehen, jedoch gibt es trotzdem Kriterien für «gut» und «schlecht». Für solche und ähnliche Probleme kann eine eigens dafür entwickelte Lösung, basierend auf Deep Learning, eine gute Alternative sein. Die Grundlage des Deep Learning sind sogenannte 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 Deep-Learning-Modells Daten benötigt. Um bei unserem Beispiel zu bleiben, werden Guetzli-Fotos von jeglichen Qualitäten mit entsprechender Bezeichnung aufgenommen. Die Daten dienen dazu, die Parameter in den neuronalen Netzwerken zu optimieren. Je vielfältiger diese Informationen sind, desto stabiler und genauer kann das Modell trainiert werden. Es gibt jedoch unabhängig von der Datenmenge keinen Garant für Erfolg.

Prozess der intelligenten Bildverarbeitung

Der Prozess der Bildverarbeitung mit Deep Learning gliedert sich in sieben Teile. Davon sind viele Teile sehr ähnlich wie in der klassischen Machine Vision ohne Deep Learning. Daher werden hier vor allem die Unterschiede hervorgehoben. In den nachfolgenden Grafiken wird eine mögliche Architektur eines Vision-Systems mit Deep Learning aufgezeigt.

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 grosse Auflösung kann indes meist verzichtet werden, da die verwendeten statistischen Modelle mit kleineren Bildern einfacher zu trainieren sind. Dies rührt daher, dass die Grösse der neuronalen Netzwerke (und entsprechend auch die Anzahl Parameter) proportional zur Bildauflösung ist. Kleinere Netzwerke ermöglichen deshalb ein stabileres Training und benötigen bei der Anwendung nur eine reduzierte Rechenleistung. Heutzutage implementieren die meisten Industriekameras 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 wichtig, 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.

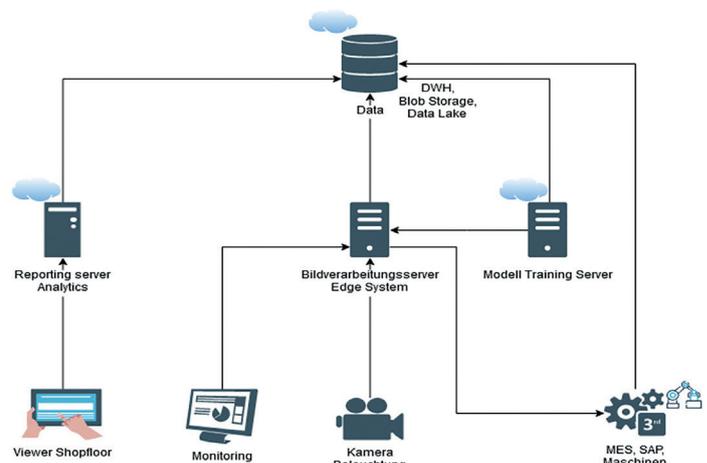
4) Bildverarbeitungs-Server / Edge-System

Wenn das Bild nun 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 Kompatibilität mit anderen Kameras ist gewährleistet, sofern die verwendete API keine spezifischen Kamerahersteller blockiert.

Die Verarbeitung der Bilder wird auf dem Server durchgeführt, wobei je nach Kamera bereits auf der Kamera selbst eine Vorverarbeitung durchgeführt werden kann. Sobald die Fotos im gewünschten Format vorhanden sind, wird das Modell auf diese angewendet. Die Bilder sowie die Resultate werden abgespeichert. Die Bilder dienen später als Trainingsdatensatz, um das Modell zu verfeinern.

Die Anforderungen an den Server sind stark abhängig von der Komplexität des Modells, der maximalen Latenzzeit sowie dem gewünschten Durchsatz. Um die Rechenzeit bei grösseren Modellen tief zu halten, kommen oft GPU (Graphics Processing Units) zum Einsatz.

Hardware Komponenten



Hardware Komponenten eines Vision Systems
Komponenten mit Wolke sind Cloud Infrastruktur



5) Modell-Training

Dem statistischen Modell und dessen Training kommt in der intelligenten Bildverarbeitung 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 z.B. von Guetzli mit entsprechenden Labels («gut» / «schlecht») gezeigt. Das Netz lernt nun aus den Bildern eine Repräsentation für ein gutes und für ein schlechtes Guetzli und berechnet aus neuen Bildern direkt die entsprechenden Labels. 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. Da das Trainieren des Modells viel Rechenleistung benötigt und nur gelegentlich ausgeführt werden muss, eignet sich dazu Cloud Computing hervorragend. Ein weiterer Vorteil der Cloud ist die Skalierbarkeit des Speichers für die Erfassung der Bildinformationen der aktuellen Bilder und der sich auch sehr gut für eine Archivierung im Cold Storage eignet.

Gewisse Cloud-Dienste wie etwa Azure Machine Learning von Microsoft unterstützen die Versionierung und den Trainings- und Deployment-Prozess der neu gewonnenen Modelle.

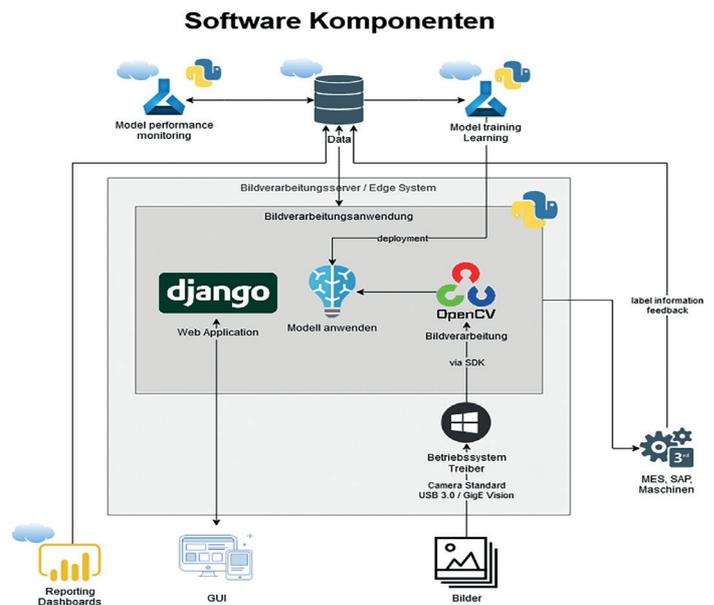
Das **Label** bezeichnet im Supervised Machine Learning (überwachten maschinellen Lernen) die Beschriftung des Datensatzes. Dies entspricht z.B. bei einem Klassifizierungsproblem der Klasse, zu dem der Datenpunkt bzw. das Bild gehört. Beispiele für Labels einer Klassifizierung in der QS könnten «ungenügend», «genügend» und «gut» sein. Ein statistisches Modell zur Generierung von Wissen aus Erfahrungswerten wird beim Deep Learning von Algorithmen aufgebaut. Es werden nicht einfach Beispiele auswendig gelernt, sondern auf Grund von Daten Muster und Gesetzmässigkeiten erkannt. Um ein Modell zu trainieren, d.h. die Parameter zu optimieren, wird jeder Datenpunkt mit einem Label benötigt. Sobald es trainiert ist, liefert das Modell zu jedem gegebenen (auch neuem) Datenpunkt das wahrscheinlichste Label.

6) GUI

Bei einer Bildverarbeitung ist es immer 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 Guetzli auch die Zuordnung zur entsprechenden Qualität wie «gut» oder «schlecht» vorliegt, um das Modell zu trainieren. Meist werden die Labels der Bilder durch einen Mitarbeitenden schon im System erfasst oder aber es muss möglich sein, dies im GUI (Graphical User Interface) zu tun. 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.



Hardware Komponenten eines Vision Systems

Die Lösung ist in Python implementiert mit Service aus der Microsoft Azure Cloud

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 laufende System ständig verbessern kann. An künstlich intelligenten Systemen wird häufig die Kritik geübt, dass sie den Menschen als Arbeitskraft wegrationieren. Dies ist bei Vision-Systemen nicht unbedingt der Fall. Denn die Arbeitslast wird zwar verringert, jedoch ist das System in vielen Fällen als Unterstützung des Menschen konzipiert. In industriellen Umfeldern kann der Produkt-Durchsatz schnell einmal eine gewisse Grenze überschreiten, bei der es schier unmöglich ist, jedes Fabrikat noch manuell zu inspizieren oder zu klassifizieren.

Gelabelte Daten

Je nach Anwendungsfall gestaltet sich die Datenerhebung etwas anders. Die Installation einer Kamera und das Sicherstellen der Bildqualität ist der erste Schritt zur Datensammlung. Jedoch ist beim Supervised Learning nebst den Bildern die Information, die im Bild sichtbar ist, also das sogenannte Label, genauso wichtig für das Training.



Im einfachsten Fall wird das Label bereits in bestehender Prozess-Software wie einem ERP oder einem Manufacturing Execution System geführt und kann von dieser Quelle direkt bezogen werden. Die manuelle Eingabe erfolgt weiterhin über die ursprüngliche Schnittstelle. Im Guetzli-Beispiel könnte es der Fall sein, dass im Prozess ohne Vision System schon für jedes Guetzli eine Eingabe in ein System dazu gemacht wird, ob es von guter Qualität ist. Zum Beispiel deshalb, weil diese Information für eine spätere Sortierung benötigt wird. In diesem Falle wäre die Möglichkeit zur Eingabe des Labels schon vor der Einführung eines Vision-Systems vorhanden.

Wo bei einer Qualitätskontrolle der Ausschuss bislang durch eine manuelle Inspektion erkannt wird, liegt eventuell keine Schnittstelle zur Eingabe dieser Information in ein System vor. Es kann zum Beispiel gut sein, dass die fehlerhaften Guetzli schon direkt bei der Inspektion von Hand herausgenommen werden und so keine Label-Informationen getrackt werden. Dann muss das Vision-System selbst eine Eingabe der Label-Information bieten.

Je nach Fall und Produktions-Setting kann die Kamera nicht gleich dort das Bild erfassen, wo auch der Ausschuss identifiziert wurde. In diesem Fall ist das Zusammenführen («Matching») von Bild und Label über ein eindeutiges Identifikationsmerkmal des einzelnen Fabrikats (z.B. einen QR-Code) zu gestalten.

Korrekte Label sind sehr wichtig. Ist die Qualität der Datengrundlage nicht gut, wird das trainierte Modell nicht die gewünschte Performance erreichen. Bei Deep-Learning-Ansätzen ist auch die Anzahl der Datenpunkte sehr wichtig. Kombinatorisch betrachtet gibt es mehr als 28'000'000 (eine 1 mit 2.5 Mio Nullen) verschiedene 1 MB grosse Bilder. Davon ist natürlich ein erheblicher Teil in der Realität nicht observierbar. Aber dennoch ist die Bandbreite an Möglichkeiten gross, weshalb eine gut funktionierende Approximation auf genügend Datenpunkten basieren muss. Konkrete Zahlen dazu, wie viele Bilder es zum Trainieren eines Modelles braucht, sind schwer zu nennen. Die Beschaffenheit des Problems (Anzahl Kategorien, Erscheinung des Fabrikates etc.) und die Architektur des neuronalen Netzes sind letztlich ausschlaggebend dafür, wie viele mit Label versehene Bilder zum Training des Netzes benötigt werden. In der Grössenordnung kommen sie meist in den Tausenden zu liegen.

Zusammenfassung

Ein Vision-System mit Deep Learning eignet sich für die visuelle Qualitätskontrolle, wenn die klassischen Machine-Vision-Methoden nicht genügen. Grund dafür sind meist fehlende gleichmässige Rahmenbedingungen und/oder fehlende Flexibilität der klassischen Modelle (neue Produkte, neue Farbvarianten, ...). Die Bausteine einer Bildverarbeitung sind neben Kamera, Beleuchtung und Trigger auch der Bildverarbeitungs-Server, das Modelltraining, ein benutzerfreundliches GUI sowie die menschliche Qualitätskontrolle. Eine hybride (Cloud/on-Premises) Architektur dieser Komponenten bietet viele Vorteile.

Wer sich für einen Deep Learning Ansatz entscheidet, setzt auf Daten. Diese gilt es in einer Qualität und einer Quantität zu halten, die es erlaubt, daraus Mehrwert zu erzielen. Die manuelle Qualitätskontrolle entfällt nicht notwendigerweise zu 100%. Mitarbeitende in der Qualitätssicherung erhalten jedoch durch ein Vision-System mit Deep Learning ein Werkzeug, mit dem sie fehlerfreier, effizienter und zielgerichteter Qualitätssicherung betreiben können.

Für Fragen und Anregungen stehen Ihnen die Autoren gerne zur Verfügung. Möchten Sie unseren Showcase zur Guetzli-Klassifizierung einmal live sehen? Dann setzen Sie sich mit uns in Verbindung.

Die Autoren



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



Jan Riedo ist Data Scientist und Data & Analytics Consultant bei IT-Logix AG. Er hält einen Master in Biomedical Engineering. Im Studium spezialisierte er sich auf intelligente Bildverarbeitung mit Deep Learning.
jriedo@it-logix.ch

IT-LOGIX AG

Schwarzenburgstrasse 11, 3007 Bern
T +41 (0)848 848 058,
F +41 (0)848 848 059

www.it-logix.ch

**LET
YOUR
DATA
SPEAK**