KI in der Videosicherheit – Chancen und Risiken – eine Standortbestimmung

Von Peter Treutler

Peter Treutler

Prokurist und Bereichsleiter IPS Intelligent Video Software bei Securiton Deutschland. Als Mitglied der Geschäftsleitung steuert er den Entwicklungsstandort in München für die Softwareprodukte IPS VideoManager und IPS VideoAnalytics und leitet den internationalen Vertrieb.

www.ips.securiton.de

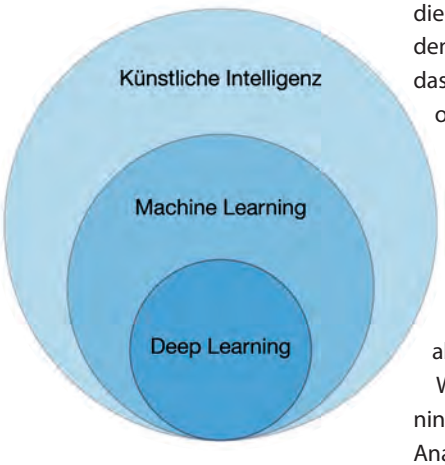

Abbildung 1: Im Artikel beschriebene Teilbereiche der künstlichen Intelligenz (insgesamt werden derzeit bis zu acht Teilbereiche unterschieden) / Bild: Securiton Deutschland

Seit einigen Jahren etabliert sich das Marketingversprechen, dass Videomanagementsysteme und zugehörige Videoanalysen durch den Einsatz von Künstlicher Intelligenz (KI) intelligenter, effektiver und somit kostengünstiger wären. Doch was bedeutet eigentlich KI? Warum erwartet man vom Einsatz KI-gestützter Systeme Vorteile? Wo liegen deren (aktuelle) Grenzen?

Der Begriff KI – Künstliche Intelligenz wurde im Jahr 1956 bei einem Treffen mehrerer Wissenschaftler im US-Bundesstaat New Hampshire geboren. Das Konzept der künstlichen Intelligenz beschäftigt die Menschheit jedoch seit vielen Jahrhunderten. Maschinen, die auf Augenhöhe mit dem Menschen Arbeit für den Menschen übernehmen können. KI ist jedoch ein sehr weit gefasster Begriff, der in aller Regel höchste Erwartungen weckt, jedoch erst einmal nichts über die Funktionalität und Qualität jedwedes Produktes aussagt. Einfach ausgedrückt werden mit dem Begriff KI Software oder Programme zusammengefasst, die Probleme alleine lösen können.

Eine spezielle Methode bzw. ein Teilbereich der KI ist das „Machine Learning" (ML). Hierunter versteht man Algorithmen, die von Daten lernen können. Im Ansatz des „überwachten Lernens" muss der Mensch hierbei die Daten und ebenso die Algorithmen vorgeben und darüber hinaus den Entscheidungsprozess überwachen (z. B. das ist ein Bild mit einem Apfel, das ist ein Bild ohne Apfel, lerne die Unterschiede und versuche Äpfel auch in anderen unbekannten Bildern zu erkennen). Beim Ansatz des „unüberwachten Lernens" muss der Computer selbst Strukturen in Daten (z. B. in Bildern) erkennen und diese sortieren. Interpretiert werden können diese sortierten Daten dann aber nur vom Menschen selbst.

Wiederum ein Teilbereich des Machine Learnings ist „Deep Learning". Hier kommen für die Analyse (sehr) großer Datensätze „künstliche neuronale Netze" (KNN, meist als NN abgekürzt) zum Einsatz. Neuronale Netze lehnen sich an die Funktionsweise des menschlichen Gehirns an. Daten werden extrahiert und analysiert und im Anschluss daran wird eine Schlussfolgerung bzw. eine Prognose erstellt (z. B. mit 90 Prozent Wahrscheinlichkeit ist ein Apfel im Bild zu sehen). Der Grundsatz des Deep Learnings unter Anwendung von neuronalen Netzen ist dabei allerdings nicht neu. Schon in den 1940er-Jahren wurde damit experimentiert. Jedoch wurden wirkliche Erfolge durch die wenig performanten Computersysteme und den fehlenden Zugang zu ausreichend großen Datensätzen ausgebremst. Durch Big Data (nahezu uneingeschränkter Zugang zu jeglichen Daten über das Internet) und entsprechend leistungsstarke Computer und Grafikkarten sind diese Hürden jedoch abgebaut. Neuronale Netze sind mittlerweile in der Lage, sehr große und unstrukturierte Datensätze (z. B. Bilder, Videos, Texte, Töne usw.) auszuwerten und Muster zu entdecken und wurden damit zum Treiber des Hypes der letzten Jahre, auch im Markt der Videosicherheitssysteme (siehe Abbildung 1).

Im Bereich der Videosicherheit haben sich KI-basierte Videomanagementsysteme und sogenannte „intelligente" Videoanalysen über die letzten Jahre sehr stark im Bereich der „Business Intelligence" (BI) etabliert. Primär marketing- oder prozesstechnisch relevante Informationen wie z. B. Altersdurchschnitt oder Geschlecht der Kunden, Kundenströme auf der Verkaufsfläche, Verweildauer von Kunden vor bestimmten Regalen usw. können mit spezialisierten „intelligenten" Videoanalysen bei entsprechender Fehlertoleranz relativ leicht bestimmt werden. Was bedeutet das? Da die Qualität eines neuronalen Netzes unmittelbar mit der Art und Menge der Trainingsdaten zusammenhängt, können vorab trainierte Systeme (z. B. durch den Hersteller während der Produktentwicklung) beim Einsatz in unterschiedlichsten Szenarien niemals eine annähernd 100-prozentige Trefferquote erzielen. Das bedeutet, dass z. B. auf die Frage „wie viele weibliche bzw. männliche Kunden waren heute

Abbildung 2: Person versteckt sich hinter einem Spiegel

im Geschäft?“ die Antwort lauten könnte, dass mit einer Wahrscheinlichkeit von 90 Prozent der Anteil männlicher Kunden bei z. B. 70 Prozent lag. Mit dieser Aussage kann man aus Marketingsicht sicherlich gut agieren und entsprechende Rückschlüsse ziehen.

Ganz anders sieht es jedoch abseits von BI-Anwendungen aus. Verwendet man Videoüberwachungssysteme im Bereich der Sicherheit, um z. B. Gebäude vor unbefugtem Eindringen zu schützen, genügt es nicht, wenn man eine Aussage wie z. B. „mit einer Wahrscheinlichkeit von 90 Prozent haben heute ca. vier Personen versucht über den Zaun zu klettern“ als Rückmeldung vom Videomanagementsystem bekommt. In diesem Umfeld werden Videoanalysen benötigt, die im besten Fall mit nahezu 100-prozentiger Genauigkeit relevante Vorfälle entdecken und melden. Aus diesem Grund haben sich Videoanalysen, die rein auf neuronalen Netzen basieren, in diesem Segment nach wie vor nicht etablieren können.

Die Herausforderung beim Einsatz von neuronalen Netzen in „freier“ Umgebung beginnen mit unerwarteten Objekten oder sehr schwer zu modellierenden Szenarien (Witterungseinflüsse, Tag/Nacht, unvorhersehbares Verhalten von Personen usw.). Um neuronale Netze zu trainieren, braucht man Unmengen von Bildern. Realistische Daten für die benötigten Einsatzszenarien sind allerdings nicht einfach zu beschaffen, da öffentlich zugängliche Bilder meist nicht den in diesen Szenarien realen Anwendungsfällen entsprechen. Personen (Einbrecher, Saboteure usw.) tarnen sich (z. B. schwarze Kleidung bei Nacht, weiße Kleidung auf Schnee usw.), verstecken sich hinter Gegenständen (siehe Abbildung 2) und bewegen sich oftmals nicht natürlich (kein aufrechter Gang, sondern z. B. Kriechen oder Robben; siehe Abbildung 3). Dies sind große Herausforderungen für gängige neuronale Netze, deren Klassifikatoren auf Basis sich natürlich bewegender Personen und selten aus dem typischen Blickwinkel einer Überwachungskamera trainiert sind.

Seit über 55 Jahren beschäftigt man sich bei Securitons Technologiemarke IPS Intelligent Video Software äußerst erfolgreich mit der Entwicklung eines ganzheitlichen und intelligenten Alarmmanagementsystems. Dieses besteht aus dem IPS VideoManager und den zugehörigen IPS VideoAnalytics und wird zum Schutz relevanter und systemkritischer Infrastrukturen wie z. B. dem Energiesektor (Energieerzeuger oder -verteiler, Petro-Chemie usw.), behördlicher Einrichtungen (z. B. Justizvollzugsanstalten) oder der produzierenden Industrie (z. B. Automotive, Lebensmittel) eingesetzt. Wegweisend wird auch hier seit mehreren Jahren daran geforscht, die etablierten und hervorragenden Videoanalyse-Algorithmen durch den Einsatz neuronaler Netze zu unterstützen und stetig weiter zu verbessern.

Initial wurde damit begonnen, die IPS Videoanalysen um Objektklassifikatoren zu erweitern und Störobjekte (z. B. Wolkenschatten, Spiegelungen, sich bewegende Büsche) auszufiltern, um so die Rate der unerwünschten Alarme stetig zu reduzieren. Im nächsten Schritt wurden NNs zur Personendetektion eingebaut, um Alarme zu verifizieren und somit die Rate der gewünschten Alarme zu verbessern. Aktuell wird daran gearbeitet, das Objekttracking (automatische Objektverfolgung mittels PTZ-Kameras über mehrere Kameras hinweg) ebenfalls durch den Einsatz von NNs zu optimieren.

Die Zukunft der neuronalen Netze liegt jedoch sehr wahrscheinlich nicht darin, sie mit immer größeren Datenmengen immer besser für bestimmte Einsatzzwecke zu trainieren, sondern darin, dass sie Anomalien erkennen. Ein NN, das über eine Art von Gedächtnis verfügt, kann selbstständig über die Zeit lernen, ob sich im analysierten Bild eine Anomalie gebildet hat (z. B. eine Person sich durch das Bild bewegt) und muss somit nicht mehr trainiert werden. Solche Algorithmen sind allerdings noch nicht einsatzbereit, wenngleich die Forschung auf Hochtouren läuft.

Abbildung 3: Diese robbende Person wird von gängigen Netzen auf Kameras nicht erkannt.