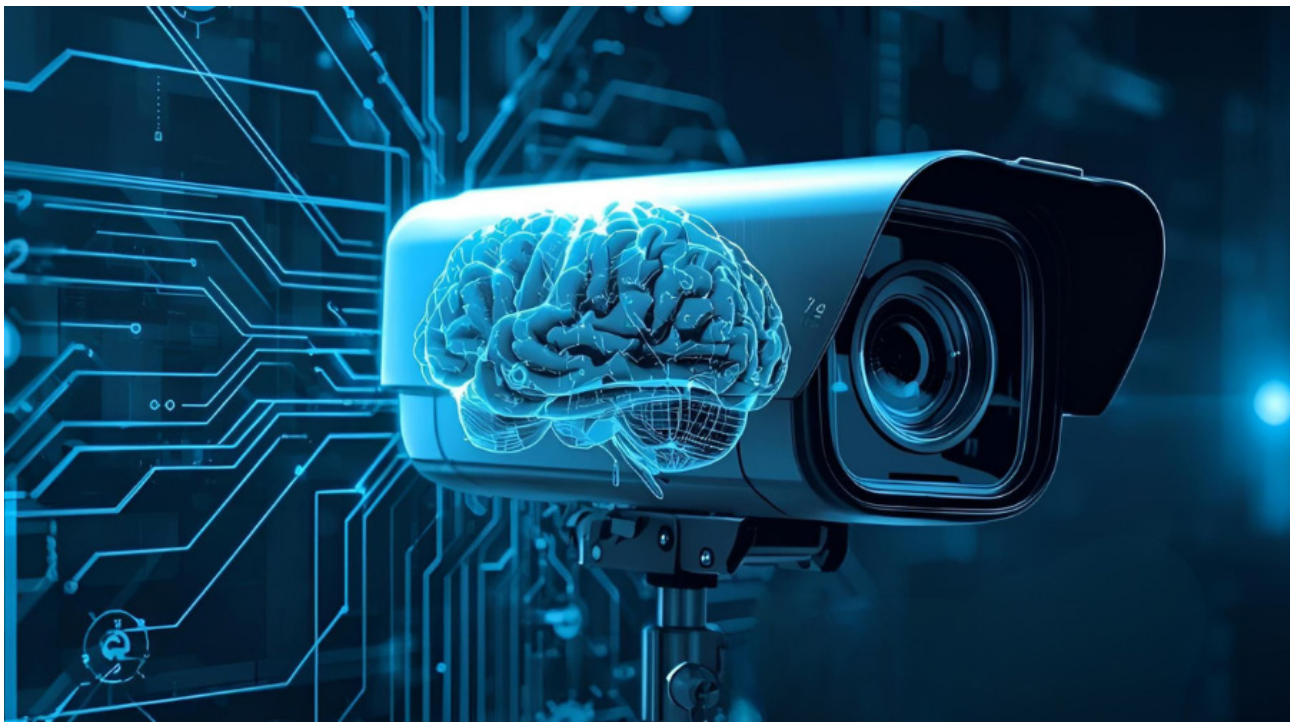
Videosicherheit

Künstliche Intelligenz (KI) in der Videosicherheitstechnik



BHE Bundesverband Sicherheitstechnik e.V.
Feldstraße 28, 66904 Brücken
Telefon: 06386 92 14-0, E-Mail: info@bhe.de,
Internet: www.bhe.de

VSS-33102-2025-10

Inhaltsverzeichnis

Vorwort

KI-Technologie ist bereits heute allgegenwärtig. Sie versorgt unsere Smartphones, bewertet und beeinflusst unsere Musikpräferenzen und leitet unsere Social-Media-Feeds. Es gibt sogar einige Lebensbereiche, wo die KI den Menschen bereits überlegen ist. Ein Beispiel sind die internationalen Börsen; wo noch vor 20 Jahren die Börsenparketts voll waren von hochbezahlten Börsenhändlern, ist heute gähnende Leere auf dem Parkett zu sehen, da die KI den Hochgeschwindigkeitshandel an allen Börsenplätzen übernommen hat. Ein weiteres Beispiel ist die Medizin, wo in der diagnostischen Radiologie die KI bei der Bildauswertung von Röntgen- Computertomographie und MRT-Auswertungen viel schneller und detaillierter als die besten Ärzte analysieren können.

„Im Allgemeinen besteht das Ziel der künstlichen Intelligenz darin, Maschinen intelligent zu machen, indem Verhalten automatisiert oder repliziert wird, sodass es einer Entität ermöglicht, angemessen und vorausschauend in ihrer Umgebung zu funktionieren", so der Informatiker und Pionier der künstlichen Intelligenz und Robotik Nils Nilsson.

Mit dem vermehrten Einsatz von KI steigt auch die Erwartung der Betreiber von Videosicherheitssystemen (VSS), um den Anwendern sowohl bei der Kriminal-Prävention als auch bei der Alarmierung und Bildauswertung die Suche in unzähligen Bildsequenzen signifikant zu erleichtern und die gewünschten Aufzeichnungen schneller zu finden.

Aber wie weit ist die KI in der Videosicherheitstechnik tatsächlich und in welchen Bereichen kann die KI die Funktionsfähigkeit eines professionellen VSS erhöhen? Kann die KI dem Anwender bei bestimmten Entscheidungen helfen oder diesen gar ersetzen? Wie sind Aussagen von Herstellern zu bewerten, wenn diese eine Trefferquote von 95% oder gar 99,9% versprechen? Wo genau liegt der Unterschied zwischen einer einfachen Bewegungserkennung (Video Motion Detection), einer Videoanalytik auf der Basis von Machine Learning und einer KI-basierten Technologie, die entweder bei der Früherkennung (Prävention) oder der forensischen Bildauswertung weitgehend vollautomatisch und korrekt arbeitet?

Oft zeigt sich in der Praxis, dass die Erwartungen an die KI viel zu hoch sind: Warum? Dies ist zum einen der oft etwas zu euphorischen Berichterstattung aus den Forschungslaboren geschuldet. Damit ist gemeint, dass Verfahren und Produkte sich noch in der Entwicklungsphase befinden und noch nicht fertig ausgereift sind. Zum anderen ist die hohe Erwartungshaltung auch den Marketingtexten einiger Marktteilnehmer geschuldet, die bereits perfekte KI-basierte Lösungen anbieten (wollen). Und letztlich werden durch die Filmstudios in spannenden Thrillern technische Möglichkeiten dargestellt, welche mit der Realität und dem tatsächlichen Stand der Technik oft sehr wenig zu tun haben.

„Bei vielen Innovationen wird außer Acht gelassen, dass neue Techniken fast immer auch eine gesellschaftliche Diskussion und Änderungen von ganz konkreten Rahmenbedingungen erfordern, bevor sie flächendeckend zum Einsatz kommen können. Das immer noch ungeklärte Dilemma beim Unfallverhalten eines autonom fahrenden Autos ist da ein fast schon klassisches Beispiel. Beim Einsatz von KI in der Videosicherheitstechnik gibt es ähnlich ungeklärte Fragen: Wieviel an Entscheidungsfreiheit erhält ein System? Welche Qualitätskriterien werden z. B. bei der Objekterkennung angesetzt? Wer ist zur Verantwortung zu ziehen, wenn z. B. eine Attacke eben gerade nicht detektiert wird, obwohl die Erwartungshaltung in der Bevölkerung möglicherweise bereits vorhanden ist? Welche Reaktionszeiten werden definiert, bis wann müssen Einsatzkräfte bei einem „KI-Alarm" vor Ort sein?

Stehen überhaupt genügend Kräfte für die potenziellen neuen Einsatz- und Rechercheoptionen zur Verfügung? Wie verhält es sich mit den vielen „False Positives", wenn z. B. über Gesichtserkennung nach einem Verdächtigen gesucht wird?

Auch mit zunehmend verbesserten Algorithmen und leistungsfähigeren Sicherheitskonzepten und -Systemen muss letztlich auch der «Mensch hinter dem System» in die Gesamtbetrachtung hinsichtlich Qualifikation und organisatorische Fähigkeiten miteinbezogen werden. Denn nur durch ein gut orchestriertes Zusammenspiel aller Faktoren ist die Einhaltung der - übrigens noch zu definierenden – Standards von Gesamtsystemen überhaupt zu gewährleisten". [1]

Obschon KI sich im Sicherheitsbereich als Technologietrend zu etablieren beginnt und bereits in einigen Branchen eingesetzt wird, befindet sie sich für viele Anwendungen im Überwachungsbereich noch in einer frühen Hype-Phase. Entsprechend sollte man vorsichtig an diese Technologie herangehen, damit diese in der Sicherheitsbranche auch zufriedenstellend funktioniert, denn es besteht immer noch das Risiko von zu hohen Erwartungen. Mit diesem Papier möchte der BHE aufklären, erläutern und dem interessierten Leser eine realistische Einschätzung zu den Möglichkeiten und Grenzen, zu den Chancen und Risiken, sowie einige Handlungsempfehlungen im Zusammenhang mit KI in der Videosicherheitstechnik geben.

Michael Meissner
Vorsitzender
BHE-Fachausschuss Video

Volker Wittchow
Stellvertreter
BHE-Fachausschuss Video

1. Begriffsdefinitionen

Künstliche Intelligenz (KI), engl. Artificial Intelligence (AI)

KI nutzt Computer und Maschinen, um Problemlösungs- und Entscheidungsfähigkeiten des menschlichen Geistes so weit wie möglich zu imitieren.

In ihrer einfachsten Form ist KI ein Bereich, der Informatik und robuste Datensätze kombiniert, um Problemlösungen zu ermöglichen. Sie umfasst auch Teilbereiche des maschinellen Lernens und des Deep Learning, die häufig in Verbindung mit KI genannt werden.

Diese Disziplinen bestehen aus KI-Algorithmen, welche darauf abzielen, Expertensysteme zu schaffen, die auf der Grundlage von Eingabedaten Vorhersagen oder Klassifizierungen vornehmen.
Schwache KI (auch Narrow AI oder Artificial Narrow Intelligence ANI genannt) ist eine KI, die auf die Ausführung bestimmter Aufgaben bzw. für dedizierte Anwendungen trainiert und ausgerichtet ist. Schwache KI ist die Grundlage für den größten Teil der KI, die uns heute umgibt. „Narrow" wäre vielleicht eine genauere Beschreibung für diese Art von KI, denn sie ist alles andere als schwach. Vielmehr ermöglicht sie einige sehr robuste Anwendungen, wie Apples Siri, Amazons Alexa, IBM Watson, die deepl.com Übersetzungsmaschine oder zunehmend auch autonome Fahrzeuge. Ebenfalls gibt es Schachcomputer, die den Menschen überlegen sind. Da diese allerdings „nur" Schach spielen können, d.h. nur für diese eine Anwendung optimiert wurden, werden sie der Schwachen KI zugeordnet.

Starke KI setzt sich aus künstlicher allgemeiner Intelligenz (AGI) und künstlicher Superintelligenz (ASI) zusammen. Künstliche allgemeine Intelligenz (AGI) oder allgemeine KI ist eine theoretische Form der KI, bei der eine Maschine über eine dem Menschen vergleichbare Intelligenz verfügen würde. Sie hätte dann ein eigenes Bewusstsein, welches in der Lage wäre, Probleme zu lösen, zu lernen und für die Zukunft zu planen.

Künstliche Superintelligenz (ASI) - auch bekannt als Superintelligenz - würde die Intelligenz und die Fähigkeiten des menschlichen Gehirns übertreffen. Starke KI ist zwar noch völlig theoretisch und ohne praktische Beispiele. Dies bedeutet jedoch nicht, dass KI-Forscher nicht auch an ihrer Entwicklung arbeiten. In der Zwischenzeit sind die besten Beispiele für ASI vielleicht in der Science-Fiction zu finden, wie z. B. HAL, der übermenschliche, schurkische Computerassistent in „2001: Odyssee im Weltraum" von Stanley Kubrick.

Machine Learning (ML)

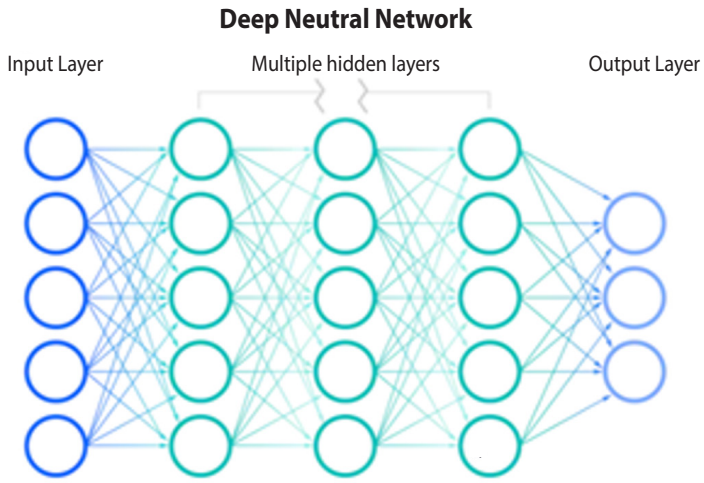
„ML ist eine Anwendung der KI, welche Systemen die Fähigkeit verleiht, automatisch zu lernen und sich aufgrund von Erfahrungen zu verbessern, ohne explizit programmiert zu werden. ML konzentriert sich auf die Entwicklung von Computerprogrammen, welche auf Daten zugreifen und diese nutzen können, um selbst zu lernen.
Der erwähnte Lernprozess beginnt mit Beobachtungen oder Daten, z. B. Beispiele, direkten Erfahrungen oder Anweisungen, um nach Mustern in den Daten zu suchen und auf der Grundlage der zuvor gelieferten Beispiele, um in Zukunft bessere Entscheidungen zu treffen. Das Hauptziel besteht darin, dass die Computer ohne menschliches Eingreifen oder Hilfe automatisch lernen und ihre Handlungen entsprechend anpassen."
[2]

Deep Learning

DL ist Teil einer breiteren Familie von Methoden des maschinellen Lernens , die auf künstlichen neuronalen Netzen mit Repräsentationslernen basieren. Das Lernen kann überwacht , teilüberwacht oder unüberwacht erfolgen.

„Deep Learning nutzt eine Reihe hierarchischer Schichten bzw. eine Hierarchie von Konzepten, um den Prozess des maschinellen Lernens durchzuführen. Die hierbei benutzten künstlichen neuronalen Netze sind ähnlich wie das menschliche Gehirn gebaut, wobei die Neuronen wie ein Netz miteinander verbunden sind. Die erste Schicht des neuronalen Netzes, die sichtbare Eingangsschicht, verarbeitet eine Rohdateneingabe, wie beispielsweise die einzelnen Pixel eines Bildes. Die Dateneingabe enthält Variablen, die der Beobachtung zugänglich sind, daher „sichtbare Schicht".

Diese erste Schicht leitet ihre Ausgaben an die nächste Schicht weiter. Diese zweite Schicht verarbeitet die Informationen der vorherigen Schicht und gibt das Ergebnis ebenfalls weiter. Die nächste Schicht nimmt die Informationen der zweiten Schicht entgegen und verarbeitet sie weiter. Diese Schichten werden als versteckte Ebenen (englisch hidden layers) bezeichnet. Die in ihnen enthaltenen Merkmale werden zunehmend abstrakt. Ihre Werte sind nicht in den Ursprungsdaten angegeben. Stattdessen muss das Modell bestimmen, welche Konzepte für die Erklärung der Beziehungen in den beobachteten Daten nützlich sind. Dies geht über alle Ebenen des künstlichen neuronalen Netzes so weiter. Das Ergebnis wird in der sichtbaren letzten Schicht ausgegeben. Hierdurch wird die gewünschte komplizierte Datenverarbeitung in eine Reihe von verschachtelten einfachen Zuordnungen unterteilt, die jeweils durch eine andere Schicht des Modells beschrieben werden." [3]



Das menschliche Gehirn wiegt im Durchschnitt 1,2 bis 1,3 kg und besteht aus bis zu 85 Mrd. Neuronen, die jeweils mit bis zu 10.000 Verbindungen zu den Nachbarneuronen ein extrem großes neuronales Netz bilden. Dies ist die Voraussetzung des Menschen zu lernen, zu schlussfolgern, abstrakt denken zu können und ebenfalls kontextuelle Zusammenhänge zu erkennen.

Video Analyse/Analytics (VA)

Mit Intelligenter VA (oder Video Content Analysis/Analytics VCA) werden alle Lösungen bezeichnet, in denen das Sicherheitssystem eine Analyse des erfassten Videomaterials vornimmt. VA sind software-basierte Analysemodule für den Betrieb auf Servern, Recordern oder in Kameras zur automatischen Erkennung sicherheitsrelevanter Objekte oder Ereignisse in Videobildern.

Sie erlauben in Echtzeit eine Objekt/-erkennung, -Verfolgung, -Identifikation, -Interpretation und-/oder eine Szenen-Interpretation. Durch ihren Einsatz verfolgen sie das Ziel, das Sicherheitspersonal zu entlasten, Datenmengen zu reduzieren und somit die Effektivität von Sicherheitssystemen zu steigern.

„VA erzeugt automatisch Beschreibungen tatsächlicher Vorgänge in einem Video (Metadaten). Diese können zum Auflisten von Personen, Fahrzeugen und anderen Objekten, die im Videostream entdeckt wurden, sowie ihres Auftretens und ihrer Bewegungen genutzt werden. Diese Informationen können dann als Handlungsgrundlage dienen, z.B. um zu entscheiden, ob Sicherheitspersonal verständigt oder eine Aufzeichnung gestartet werden sollte." [4]

Metadaten (im Zusammenhang mit VA)

Metadaten sind im Wesentlichen Daten über diese Daten. Abhängig von ihren Einstellungen können die Metadaten die Uhrzeit und das Datum, an dem das Video erstellt wurde, Brennweite und Verschlusszeit, aber auch die im Video analysierten Objekte und Eigenschaften wie Größe, Farbe, Bewegungsvektoren und noch vieles mehr mitteilen. Durch die Verwendung der Daten und Metadaten steigt die Effizienz bei der Analyse und damit auch deren Zuverlässigkeit.

Um möglichst valide Metadaten von Objekten in unterschiedlichen Szenen und Ansichten zu erhalten, benötigt es bei gängigen Methoden neben sehr vielen Trainingseinheiten auch einen großen Pool bereits klassifizierter Szenen und Objekte in der Szene. In den letzten Jahren gab es große Fortschritte bei alternativen Methoden, mit denen performante Modelle mit nur sehr wenig klassifizierten Szenen erzeugt werden können. Diese Methoden benötigen sehr viele nicht klassifizierte Szenen für das Training, und einige wenige klassifizierte Szenen reichen dann aus (für gewöhnlich weniger als 100), um das Modell für die individuelle Anwendung anzupassen.

GPU (Grafikprozessor oder Graphics Processing Unit)

GPUs werden häufig als Grafikkarten oder Videokarten bezeichnet. Jeder PC verwendet einen Grafikprozessor, um Bilder, Videos und 2D- oder 3D-Animationen für seine Anzeige wiederzugeben. GPUs wurden dediziert entwickelt um grafikrelevante Daten – auch Videostreams möglichst effizient darzustellen. Sie führen schnelle mathematische Berechnungen durch und geben dadurch die CPU für andere Aufgaben frei.
Während eine CPU einige wenige Kerne verwendet, welche auf sequenzielle, serielle Verarbeitungen ausgerichtet sind, verfügt eine GPU über tausende kleinerer Kerne, die für Multitasking ausgelegt sind.

Es gibt zwei verschiedene Arten von GPUs:

- Integrierte GPUs befinden sich auf der CPU eines PCs und teilen sich den Speicher mit dem Prozessor der CPU
- Diskrete GPUs befinden sich auf einer eigenen Karte und verwenden ihren eigenen Videospeicher (VRAM), damit der PC seinen Arbeitsspeicher nicht für die grafikrelevante Operationen verwenden muss.

Die Rechenkapazität wird in TeraFLOPS (Floating Point Operations Per Second) angegeben.

TPU (Tensor Processing Unit) und NPU (Neural Processing Unit)

TPUs (Tensor Processing Units) und NPUs (Neural Processing Units) sind spezialisierte Chips für KI-Beschleunigung. Während GPUs ursprünglich für Grafikberechnungen optimiert waren und erst später für KI-Anwendungen genutzt wurden, sind TPUs und NPUs von Anfang an mit dem Ziel konstruiert worden, neuronale Netze effizienter zu berechnen.

- TPUs wurden von Google entwickelt und sind auf die Verarbeitung sogenannter Tensoroperationen optimiert, die eine zentrale Rolle im Deep Learning spielen. Sie werden überwiegend in Rechenzentren eingesetzt, um große KI-Modelle zu trainieren oder zur Inferenz (Ausführung) bereitzustellen.
- NPUs sind speziell darauf ausgelegt, künstliche neuronale Netze mit möglichst geringem Energieverbrauch und hoher Geschwindigkeit zu verarbeiten. Sie kommen vor allem in mobilen Endgeräten, Edge-Computing-Systemen oder autonomen Anwendungen zum Einsatz.

Kamera-basierte «Edge basierte» Analyse

Algorithmen für VA, welche direkt in der Kamera (on the edge) laufen. Höherwertige Kameras verwenden hierzu dedizierte, für Videoanalyse optimierte GPUs und NPUs. Da intelligente Kameras durch die eingebaute VA eine Datenverarbeitung am «Rande des Netzwerks» ermöglichen, wird dies auch als Edge Computing bezeichnet.

On-Premises

„On-Premises" bedeutet im Deutschen so viel wie „in den eigenen Räumlichkeiten" oder „vor Ort". Diese Definition von On-Premises bezieht sich auf die Nutzung unternehmenseigener Server und der eigenen IT-Umgebung. Da der Nutzer die Software im eigenen Rechenzentrum auf eigener oder gemieteter Hardware betreibt, spricht man auch von „Inhouse". Im Gegensatz zu Cloud-Computing erhalten Kunden bei On-Premises die vollständige Kontrolle über die Daten und übernehmen auch alle damit verbundenen Risiken in eigener Verantwortung." [5]

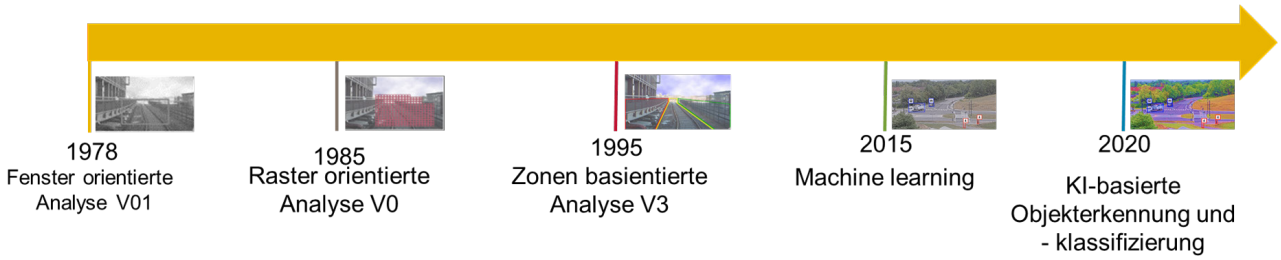
Cloud-Systeme

„Bei der Cloud handelt es sich um einen serverbasierten Datenverarbeitungsprozess, der sich um große, zentralisierte Server in Datenzentren dreht, welche von Drittanbietern verwaltet und gewartet werden. Nachdem die Daten auf einem Endgerät, z. B. einer Kamera, erzeugt wurden, werden sie zur Speicherung und Verarbeitung an den zentralen Server weitergeleitet.
Die Cloud-Analyse gilt als flexibel und kann je nach Gesamtnutzung, Benutzeranforderungen und Größe der Anforderungen effektiv und rasch nach oben oder unten skaliert werden." [6]

Hybride Systeme

Sowohl die Cloud-basierten als auch die On-Premise-Ansätze haben jeweils Vor- und Nachteile und somit kann keine der erwähnten Lösungen die andere vollständig substituieren. Hybride Systeme kombinieren daher beide Systemarten und bieten je nach Anwendung optimierte Analyseergebnisse.

2. Historische Entwicklung



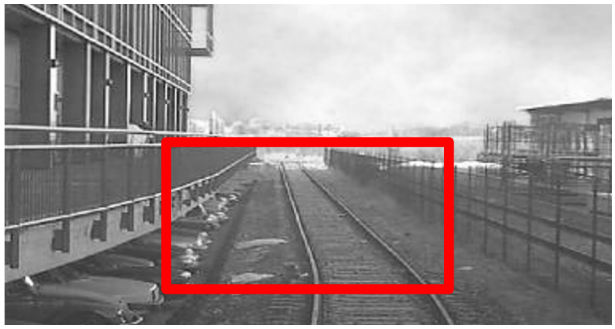
Videoanalyse (erste einfache Sensoren)

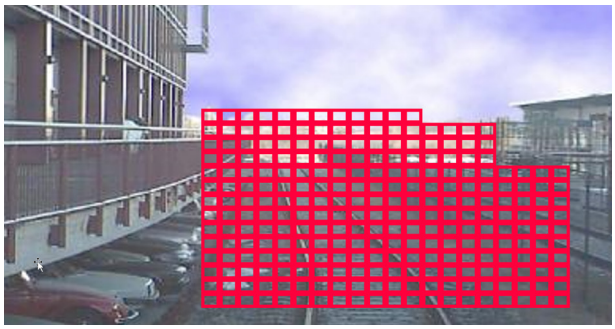
Die ersten Entwicklungen hatten den Fokus, Helligkeitsabweichungen im Bild zu erkennen und bei Veränderungen einen Alarm auszulösen mittels pixelbasierter Bildpunktprüfung zwischen einzelnen Bildern.

Da die Rechenleistungen in den frühen 70er Jahren ein Bruchteil der heutigen gewesen sind, beschränkte man sich daher erst einmal auf ein „Fenster" in der Szene, welches auf Helligkeitsveränderung „überwacht" wurde. Derartige Lösungen wurden durch die Begriffe «activity detection» oder «motion detection» geprägt.

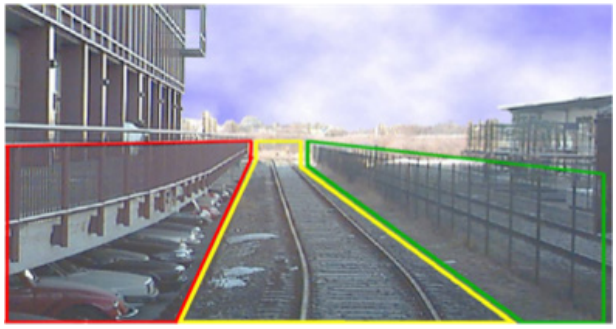
Aufgrund von umweltbedingten Einflüssen, wie z. B. durch Wind verursachte Bewegungen von Sträuchern, Bäumen, Laub, als auch Regen, Schnee etc. wurden sehr viele Falschalarme ausgelöst.

Im Laufe der Zeit wurde versucht, die Anzahl an Falschalarmen durch Nutzung von verschiedenen Software-Filtern zu reduzieren. Mit „Empfindlichkeitsreglern“ (Regelschieber) wurde Einfluss genommen z. B. auf die Anzahl an geänderten Bildpunkten oder aber an der zeitlichen Länge (Anzahl an Einzelbildern) um einen Alarm/Ereignis auszulösen.





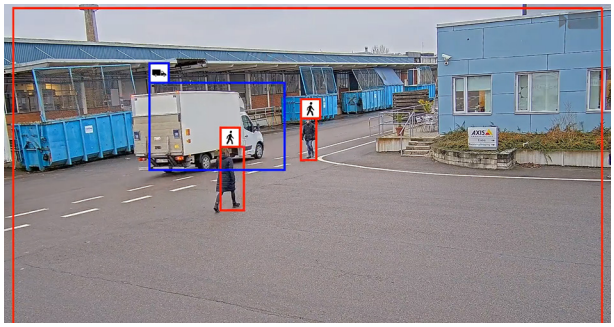
Aufgrund der durch wachsenden Terrorismus und organisierter Kriminalität angespannten Lage wurde dem Bereich der Videoanalyse immer mehr Aufmerksamkeit geschenkt. Aus einem „Fenster“ wurden zunächst mehrere und die ersten Abhängigkeiten von verschiedenen Fenstern untereinander geschaffen. Daraus entstanden rasterorientierte Systeme.



Mit zunehmender Rechenleistung ließen sich auch immer bessere Algorithmen integrieren und daher war es nur ein kleiner Schritt, um aus einem 2D-Videobild eine 3-dimensionale Szene zu konstruieren. Dadurch ist es gelungen, die Tiefe und auch die Größenabhängigkeit mit der zunehmenden Entfernung zu berücksichtigen. Dies ermöglichte es nun, nicht nur Objekte zu erkennen, sondern ihnen auch noch weitere Objekt – Attribute in einer 4D-Betrachtung zu integrieren. Die erwähnten Attribute können neben der Geschwindigkeit auch Größe, Richtung, ein Verhalten oder eine Strecke, welche zurückgelegt werden muss, sein. Dies sind nur ein paar Möglichkeiten von Attributen.

VMDs, Stand heute

Aufgrund der enormen Entwicklung im Bereich der Prozessoren für Computer, Spielkonsolen und Smartphones und der damit entwickelten Grafikprozessoren, die auf die Berechnung von Grafiken spezialisiert und optimiert sind, steht nun der nächste Durchbruch im Bereich der Videoanalyse an. Hier ist die Kombination der 3D-Analyse mit und ohne KI zu sehen. Der Vorteil einer KI-basierten Analyse liegt in der Bestimmung der „Objekte von Interesse“ und deren Verfolgung.



Moderne KI-basierte VMDs arbeiten nicht mehr auf Basis pixelbasierter Bildpunktprüfung, sondern „lernen“ Dinge im Bild zu erkennen, wie z. B. Menschen, Fahrzeuge, Tiere. Also eine Objekterkennung mithilfe von computergenerierten Algorithmen, die auf Trainingsdaten bzw. Bildern basieren.
Mittels der genutzten neuronalen Netze lernt die KI-basierte Analyse, den Algorithmus zu verbessern. Dies hat u. a. den Vorteil, dass z. B. weniger Falschalarme ausgelöst werden, da sich durch Wind bewegende Bäume, Äste oder Regen nicht mehr als Ereignis gewertet werden.

Künftige Entwicklungen

Da wir nach dem Mooreschen Gesetz, also nach der Verdopplung der „Rechenleistung" alle circa 18 Monate auch weiterhin mit leistungsstärkeren Prozessoren rechnen können, wird es nur eine Frage der Zeit sein, dass wir immer validere „Objekte von Interesse" bestimmen können. Hierzu können dann auch Eigenschaften wie der Gang oder aber die Körperhaltung gehören.

Dies ist aber nur ein Aspekt. Ein weiterer sehr wichtiger Punkt hinsichtlich der Einhaltung der DSGVO ist, dass auch „Objekte von Interesse" anonymisiert werden können, das heißt, dass Personen zum Beispiel nicht mehr als Person zu erkennen sind, sondern als anonymisierte Avatare.

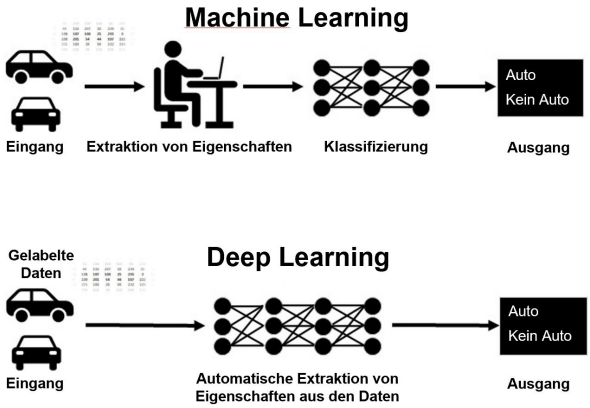
Einschränkungen anhand praktischer Beispiele

Im Bereich der Absicherung von größeren Außenbereichen, z. B. Perimetersicherung, Geländeabsicherung Grenzsicherungen o. ä. wird eine auf Menschen erkennende KI ein gutes Detektionsergebnis immer dann erzielen, wenn der Mensch mit seinen als Mensch-typischen Merkmalen und Bewegungen vom Algorithmus erkannt wird. Allerdings muss man wissen, dass es mit hoher Wahrscheinlichkeit zu fehlenden Detektionen kommen wird, wenn ein Mensch sich durch entsprechende Tarnung (Tarnfarbe der Kleidung passend zur Natur) und atypischen Bewegungen (z.B. langsam durch das Bild rollend) auf ein Angriffsziel zubewegt; weil der Algorithmus diese Art der Bewegung „noch nicht" gelernt hat. Eine besondere Herausforderung in diesem Zusammenhang spielt auch der Blickwinkel der Kamera (in dieser Anwendung meistens eher weitwinklig), weil „rollende Menschen" im vorderen Erfassungsbereich der Kamera relativ groß dargestellt werden, während ein paar Meter dahinter die gleiche Person relativ klein dargestellt wird.

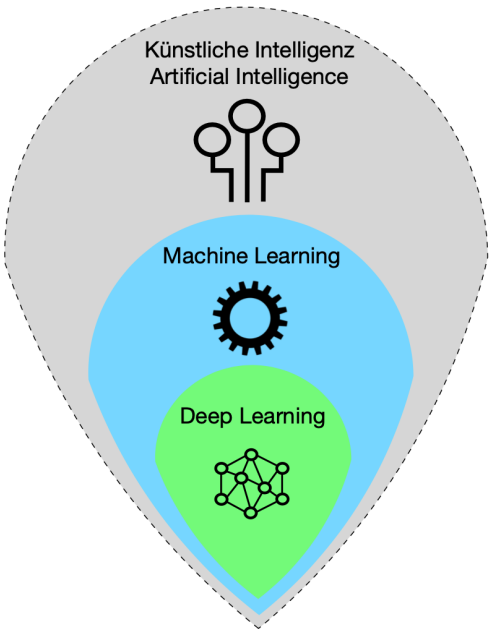
Ebenso ist es wichtig zu verstehen, dass eine KI-basierte Videoanalyse keine Kontext-abhängige Bewertung der Bilder vornehmen kann. Wenn z. B. ein Mensch erkennbar schnell durch das Bild läuft, kann die KI nicht erkennen, warum diese Person schnell läuft. Versucht die Person vor einer direkten Gefahr zu fliehen oder will diese nur schnell den Bus an der Haltestelle erreichen. Menschen können dies relativ schnell erkennen. Eine KI kann dies nicht analysieren, solange sie nicht speziell auf diese Fähigkeit trainiert wurde.

3. Grundlagen der KI in der Videosicherheit

Damit Videoinformationen von Maschinen (egal ob in der Kamera oder auf einem Server) überprüft und analysiert werden können, müssen sie in Daten umgewandelt werden. Dies geschieht durch eine Datenabstraktion, die sich auf die Reduzierung einer Datenmenge auf eine vereinfachte Darstellung des Ganzen bezieht. Eine mit Analytics ausgestattete Sicherheitskamera erkennt eine Person nicht auf dieselbe Weise wie Menschen, sondern versteht die wesentlichen Merkmale der Person als Daten.
Und diese Analytics wandelt nicht nur Videoinformationen in Daten um (was in der Szenenanalyse weitgehend genutzt wird), sondern erstellt auch Metadaten.



Typische Eigenschaften				
	Trainings-datensätze	Festlegung der Eigenschaften	Benötigte Rechenleistung	Trainingszeit
Machine Learning	Wenige (1.000+)	Ja	Hoch	Kurz
Deep Learning	Viele 1.000.000+)	Nein	Sehr hoch	Lang



Durch Machine Learning (ML) können Computer im Gegensatz zur traditionellen Programmierung ihre eigene Logik erstellen, um Vorhersagen zu treffen und um entsprechende Schlussfolgerungen zu ziehen.

Wie das links dargestellte Bild zeigt, ist Deep Learning eine Unterkategorie von Machine Learning und damit eine fortgeschrittene Form der KI.

„Die Begriffe AI, Machine Learning und Deep Learning werden oft miteinander vermengt und synonym verwendet. Doch es gibt Unterschiede. Grob gesagt bildet Machine Learning eine Untermenge von AI, und Deep Learning entspricht einem Teilbereich von ML. Von Artificial Intelligence spricht man, wenn Maschinen Aufgaben übernehmen, die menschliche Intelligenz imitieren, indem sie beispielsweise planen, lernen und Probleme lösen. Das betrifft aber auch das Verstehen von Sprache oder das Erkennen von Objekten, Bildern oder Geräuschen." [7]

Die Möglichkeiten, die die „künstliche Intelligenz" in der Videosicherheitstechnik eröffnet, ermöglichen eine immer bessere Erkennung und Verfolgung von Objekten von Interesse, wie beispielsweise Autos, Personen oder Fahrräder, in der Szene. Durch die intelligente Verknüpfung der KI mit den „konventionellen" Analyse Algorithmen wird es nun wesentlich leichter und dadurch auch sicherer, zum Beispiel Personen in der Szene zu erkennen und zu verfolgen. Dieses sicherere Erkennen und die sicherere Verfolgung eröffnet nun weitere Möglichkeiten, die eine drastische Reduzierung der unerwünschten Meldungen zu Folge hat. Zurzeit werden bei der professionellen Videoanalyse im Freigelände, die die Königsklasse der Video Content Analysis (VCA) ist, beide Techniken genutzt.

4. Trainingsmethoden von Videoanalyse

Maschinelle Algorithmen für ML und DL ermitteln anhand vieler Beispiele die Beziehungen zwischen den Eingabedaten und deren Schlussfolgerungen. Der Vorteil gegenüber der traditionellen Programmierung besteht in der Anzahl der Beispiele, welche ein Computer durcharbeiten kann, ohne den Fokus zu verlieren und in denen er relevante Eigenschaften finden kann.

Der Hauptunterschied zwischen Technologien des ML und DL besteht in der Datenmenge, welche für das Training eingesetzt werden muss. Während es bei ML normalerweise um ein paar tausend Bilder geht, werden bei DL Millionen Datensätze verwendet.

Je mehr Daten beigezogen werden, desto mehr Szenarien können trainiert werden. Will man beispielsweise Objekte erkennen, können mit steigender Menge der Eingabedaten mehr Objektklassen oder Variationen derselben Objekte erkannt werden. Dies wird benötigt um zum Beispiel verschiedene Körperhaltungen, oder auch teilweise verdeckte Menschen zu erkennen.

DL verwendet neuronales Netzwerk-Modelle. Ein untrainiertes DL- Modell hat noch nichts gelernt und ist noch „leer". Um das Training eines Deep Learning Algorithmus zu verdeutlichen, kann man dies ähnlich wie bei der Lernphase eines kleinen Kindes betrachten: Das Kleinkind hat ein Gehirn, welches erst mit Wissen gefüllt werden muss. Dessen Training beginnt z. B., als würde man mit dem Kind im Kindergarten ein Bilderbuch betrachten und ihm beibringen „dieses Objekt ist ein Auto, ein anderes Objekt ist ein Fahrrad" etc.

Der nächste Schritt ist die Optimierung des Modells mit einem speziellen Training. Dazu werden weitere Objekteigenschaften angelegt, wie zum Beispiel, dass der Mensch rennt, ein Auto fährt, oder dass dies ein spezifisches Auto ist wie z.B. ein roter Lieferwagen von Volkswagen o. ä. Das Kind ist jetzt in der Grundschule angekommen.
Und wenn das Training abgeschlossen ist, ist aus dem Kind ein Erwachsener geworden, der sein Studium abgeschlossen hat und im Vergleich zu AI ist das Modell / der Algorithmus, selbst in der Lage Schlussfolgerungen zu ziehen und weiter zu lernen.

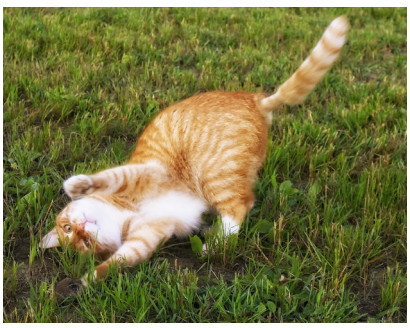
Die KI soll im Idealfall die menschliche Intelligenz nachbilden. Doch, obschon man dies gerne mit dem Begriff KI verbindet, besitzt ein DL-Algorithmus kein Bewusstsein oder keine allgemeine Intelligenz wie ein Mensch.

Um den vorherigen Vergleich eines Kindes nochmals heranzuziehen, wird dieses begreifen, dass das erste Bild der nachfolgenden Sequenz eine Katze ist. Dies kann die KI ebenfalls. Aber stellt das zweite Bild ebenfalls eine Katze dar und das Dritte ebenfalls? Oder handelt es sich vielleicht um ein Auto bei Nacht?

Bild 1a

Katze

Bild 2a



Ebenfalls Katze?

Bild 3a

Katze, Hund oder Auto?

Ein Kleinkind lernt bereits kurz nach seiner Geburt haptische, akustische und visuelle Eindrücke zu interpretieren und i.d.R. bis zu seinem Tod Neues zu lernen. Im folgenden Beispiel ist es für ein menschliches Gehirn ohne Probleme möglich zu erkennen, dass es sich bei Bild 2 und 4 um einen Schneemann handelt, wobei Letzterer atypisch «auf dem Kopf» steht. Bild 1a und 3a illustriert einen als Schneemann verkleideten Menschen.

Bild 1a

Bild 2a

Bild 3a

Bild 4a

Für einen KI gestützten Algorithmus ist diese Aufgabe ungleich schwieriger zu lösen. Nur wenn die Trainingssequenzen entsprechende Bilder enthalten, ist diese Aufgabe für ihn lösbar, wobei atypische Szenen im normalen Umfeld eben kaum oder gar nicht vorkommen.

Die Möglichkeiten des eingesetzten DL-Modell hängen davon ab, wofür dieses trainiert wurde. Das Training wird während den Trainingssequenzen i. d. R. durch Menschen sorgfältig überwacht, da es sonst leicht verzerrt werden kann.

Dies könnte z. B. entstehen, wenn das DL-Modell mit vielen Bildern von Zebras und nur einigen wenigen Pferden trainiert wird. Als Resultat wird die Schlussfolgerung des Algorithmus vermutlich ebenfalls eher ein Zebra erkennen, selbst dann, wenn es sich tatsächlich um ein Pferd handelt. Auch wenn die Pferde auf den Trainingsbildern auf Gras laufen wird es weniger wahrscheinlich sein, dass das finale Modell ein Pferd auf einer Straße erkennen wird.

Menschen sind derzeit noch einiges schlauer als DL-Algorithmen. Wir können jedoch leicht die Assoziation übersehen, welche das neuronale DL-Netzwerk macht und welche für uns unlogisch erscheinen, wie die vorher erwähnte Assoziation zwischen Pferd und Gras. Um belastbare Algorithmen zu erlangen, werden demnach unzählige relevante und voreingenommene Trainingsdaten benötigt.

Im Videobereich werden üblicherweise begleitete Trainingsmethoden (supervised training) eingesetzt.
Dazu werden 100.000de bis Millionen von vorab klassifizierten Bildern benötigt, um ein einigermaßen breit gefächertes und akkurates Ergebnis zu erzielen. Der dabei entstehende Trainingsaufwand ist daher i. d. R. sehr hoch, entsteht allerdings mehrheitlich initial (meistens einmalig). Dem Algorithmus wird dabei antrainiert, was man finden möchte und aber auch das Gegenteil (was man nicht möchte).

In den letzten Jahren hat sich eine unbegleitete Trainingsmethode namens Self-Supervised Learning (SSL) hervorgehoben. Bei SSL werden Modelle nicht mit expliziten Labels trainiert, sondern lernen, intrinsische Strukturen und Muster in den Daten selbst zu erkennen. Statt eines Menschen, der jedes Bild klassifiziert, erzeugt das System automatisch Lernaufgaben aus den Rohdaten.

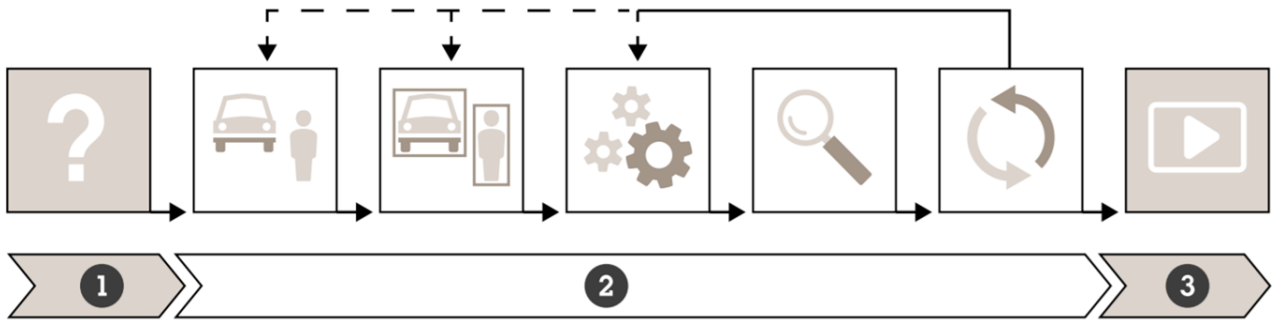
In einem zweiten Schritt kann das Modell dann mit deutlich weniger Rechen- und Datenaufwand mithilfe von supervised training oder mit einem simplen Klassifikator auf spezifische Anwendungen angepasst wer-

den. Diese Methode erzeugt mittlerweile ähnlich gute Ergebnisse wie supervised training und eröffnet neue Möglichkeiten für effiziente Videoanalyse, insbesondere bei kleinen und ungewöhnlichen Datensätzen.

Das Training/die Klassifizierung wird bei beiden Trainingsmethoden nicht am Zielgerät, sondern mit dezidierten Rechnern sehr großer Leistung, oder Cloud-basiert umgesetzt. Als Ergebnis des Trainingsvorgangs entsteht ein Modell. Für Videoanalyse beschreibt dieses die nötigen Parameter für die Ausführung auf CNNs Convolutional Neuronal Networks (CNN). Für andere Deep Learning Lösungen gelangen auch andere Trainings-Methoden zum Einsatz.

Das Modell wird nach Abschluss des Trainings auf das Zielgerät geladen (z.B. Deep Learning fähige Kamera) und dort auf einer GPU oder NPU ausgeführt.

1. Vorbereitung: Festlegung des Anwendungszwecks.
2. Training: Sammeln von Trainingsdaten. Versehen der Daten mit Anmerkungen (Annotieren). Trainieren des Modells. Testen des Modells. Ist die Qualität nicht wie erwartet, werden die vorangegangenen iterativen Verbesserungsschritte wiederholt.
3. Fine-Tuning: Nach dem initialen Training kann das Modell gezielt mit für die Anwendung spezifischen Daten nachtrainiert werden, um die Genauigkeit und Robustheit für die konkrete Aufgabe zu erhöhen.
4. Deployment: Installation und Verwendung der fertigen Anwendung



Für Videoanalyse wird zusätzlich noch mehr benötigt (Tracking, Rules Engine, Event handler etc.). Dies verhält sich ähnlich wie bei Deep Learning in einem iPhone: Apple trainiert die Modelle mit anonymen Daten und synthetischen Daten, dessen Modell z. B. für die Fotoerkennung in das iPhone portiert werden. Verbesserte Modelle auf Basis erweiterter Trainingssequenzen können bei Updates wieder in das Smartphone geladen werden, um weitere Möglichkeiten z. B. bei der Objekt-Klassifizierung, oder verbesserte Genauigkeit zu erhalten.

Was wird klassifiziert:

Menschen

Klassifizierung verschiedener Objekte wie z. B. Mensch, Tier, Fahrzeug, Gegenstände etc.; Menschen mit unterschiedlicher ethnischer Herkunft, erschwert durch unterschiedliche Kamera-Winkel, Kamera-Neigungen, Belichtungen etc. z. B. eine Million Bilder. Je mehr unterschiedliche Datensätze zum Training verwendet werden, desto besser kann das System in die Lage versetzt werden, auch nur (visuell sichtbare) Teile eines Menschen wie z. B. eine Hand oder eine Körperhälfte zu erkennen und einem Menschen zugehörig zu klassifizieren.

Fahrzeuge

Dito, jedoch in der Regel mit Subklassen wie PKW, Bus, LKW, Fahrrad, Motorrad etc.

Aus der erwähnten Aufgabenstellung der Klassifizierung unterschiedlicher Objekte ergibt sich: Je mehr Klassen, umso mehr Training ist erforderlich.

5. Plattformen-Übersicht

Hochwertige Kameras verfügen heutzutage über leistungsfähigere Prozessoren. Dies erlaubt oftmals einen Parallelbetrieb der Videoanalyse und der Encodierung der Videodaten durch einen einzigen Prozessor. Eine weitere Variante ist der Einsatz eines zusätzlichen, dedizierten Prozessors für die Videoanalyse in der Kamera. Derartige Lösungen nennt man auch kamerabasierte Analyse **(Edge-basierte Analyse)**.
Die Vorteile dieser Betriebsart liegt in der Echtzeitverarbeitung, niedrigeren Kosten und einer verringerten Komplexität des Gesamtsystems. Dabei werden oft weniger Bandbreite benötigt und zudem der Datenschutz verbessert, weil nur relevante Videos über das Netzwerk übertragen werden.

Server-basierte Videoanalyse ist momentan wohl noch die verbreitetste Lösung, bei welcher ebenfalls zwei unterschiedliche Systemarchitekturen realisiert werden können. Entweder wird die Videoanalyse durch eine leistungsfähige CPU durchgeführt oder die Algorithmen benutzen die Grafikkarte, welche gegenüber einer konventionellen Computer CPU eine signifikant höhere Performance für rechenintensive Analysen von Videostreams bereitstellen kann. Diese Variante ist sehr flexibel bezüglich der Zuordnung der notwendigen CPU Performance und pro Videostream können mehrere Analysemodule parallel betrieben werden.
Je nach Bedarf besteht die Möglichkeit eine hybride Architektur einzusetzen, welche eine kamera-basierte mit einer server-basierten Analyse kombiniert. Denkbare Szenarien sind auf der Kamera laufende Algorithmen, welche Personen zählen und die daraus entstehenden Metadaten auf dem nachgelagerten Video-Managementsystem mit weiteren Analysen verarbeitet und für die eine nachträgliche, forensische Auswertung aufbereitet und bereitstellt.

Cloud-basierte Analyse
Da Videoanalyse per se rechenintensive Vorgänge bedeuten, bieten einige Anbieter zunehmend auch Videoanalyse-Lösungen in der Cloud an.
Zur Analyse können hierbei folgende Streams des lokalen Systems übertragen werden:

- Übertragung von Videostreams, Analyse in der Cloud
- Übertragung von reinen Metadaten, welche in der Cloud analysiert und ausgewertet werden
- Übertragung beider Elemente zur erweiterten Analyse in der Cloud

Anders als typische On-Premise Systeme, bei denen der Nutzer die Gesamtlösung in der Regel installieren lässt und somit kauft, werden Cloud-basierte Ansätze in der Regel als Mietmodell angeboten. Der Kunde „abonniert“ und bezahlt hierbei nur für eine zeitlich definierte Leistung.

Hybride Architektur
In einer hybriden Mischform werden die beiden oben genannten Lösungsansätze einer On-Premise-, als auch eines Cloud-basierten Videoanalyse-Systems umgesetzt.
Bei Edge-Geräten wie IP-basierten Kameras, Audio-Recordern und anderen Sensoren muss zwangsläufig ein Gleichgewicht zwischen Cloud- und On-Premise-Optionen gefunden werden. Die Vor-Ort-Option wird für die Verarbeitung zeitkritischer Daten bevorzugt, während das Cloud-Computing für die Verarbeitung von Daten verwendet wird, die nicht zeitabhängig sind.

Die Umsetzung hybrider Architekturen kann daher auch ökonomisch motiviert sein, indem On-Premise nur Teile aller gewünschten Videoanalysen laufen und man zusätzlich in der Cloud mit leistungsfähigen Servern weitere Analyse-Features bereitstellen und skalieren kann.

Bei bestimmten Anforderungen kann daher eine Kombination aus Edge- und Cloud- oder einer Hybrid-Computing-Infrastruktur das beste Ergebnis liefern.

6. Typische Anwendungsfälle von KI in Videosicherheitsanlagen

Die nachfolgend aufgeführte Liste ist keinesfalls abschließend, sondern greift einige exemplarische Anwendungen heraus, bei denen bereits KI-basierte Videoanalysen eingesetzt werden.

Safety Applikationen
Detektion von Feuer und Rauch. Hierbei dient die Videoanalyse zur Brand-Früherkennung. Dabei erkennen die Algorithmen Flammen und-/oder Rauch direkt an der Entstehungsquelle. Sie kann daher Brände häufig schneller identifizieren als konventionelle Melder an einer Decke. Durch das Fehlen entsprechender Normen zur eindeutigen Qualitätsbeurteilung werden derartige Lösungen i. d. R. als freiwillige Systeme eingesetzt und ersetzen keine herkömmliche Brandmeldesysteme.

Physical Security / Perimeter-Sicherheit
Typische Anwendungsfälle sind kritische Infrastrukturen, JVAs, Industrie/Produktionsbetriebe, Logistikzentren, private Anwesen etc. Es geht hierbei in der Regel um den Perimeterschutz, also die Sicherung des Grundstücks bzw. des Areals bereits ab der Außengrenze. Im Kern geht es darum, dass nur befugte Personen das Gelände betreten – oder verlassen dürfen. Die Absichten dahinter können unterschiedlich sein, z. B. Einbruch- und Diebstahlschutz, Schutz vor neugierigen Blicken, Angst vor Sabotage und Vandalismus, oder die Detektion eines Ausbruchsversuchs in JVAs.

Die KI ermöglicht durch eine Steigerung der Präzision/Genauigkeit der Detektion eine Verbesserung der Ergebnisse bei gleichzeitig deutlicher Reduktion möglicher Falschalarme infolge auftretender Störgrößen im Außenbereich wie Schattenwurf, sich im Wind bewegende Pflanzen, Schneefall, Kleintiere etc.

Autokennzeichen-Erkennung
Typische Anwendungsfälle sind das automatische Öffnen von Schranken bei Zufahrten von registrierten Nummernschildern, automatisierte Ticketing- und Rechnungsstellung, Protokollierung der Verweildauer von Fahrzeugen für Business Intelligence Auswertungen, Steuerung von Parkleitsystemen für treue, wiederkehrende Kunden und VIPs etc., Verfolgung von Benzindiebstahl in Tankstellen etc., Erhebung von Road Pricing in Städten oder Autobahn-Maut, Identifizieren gesuchter Fahrzeuge bei Grenzübertritt etc.

Gesichts-Erkennung
Typische Anwendungsfälle sind das biometrische Identifizieren von Gesichtern von Personen, die auf einer schwarzen Liste (black list) stehen, oder eine berührungslose Zutrittskontrolle, um den Zugang auf physische Standorte wie Gebäude, Büros, Rechenzentren und andere Sicherheitsbereiche zu kontrollieren (white list). Die Gesichtsbiometrie wird hierbei verwendet, um einen Benutzer aus Zugriffskontrolllisten zu identifizieren und seine Identität zu überprüfen.

Gesichtserkennungssysteme können Gesichter in Echtzeit mit Datenbanken bekannter Straftäter abgleichen und helfen so bei der Identifizierung Verdächtiger. Es gibt KI-Systeme, die Gesichter automatisch erkennen, um sie dann zu anonymisieren (z. B. Verpixelung), damit Datenschutzauflagen erfüllt werden können.

Unterstützt durch Auflagen, welche rund um die Pandemie Covid-19 entstanden, erfuhren Lösungen zur Analyse von Maskenträgern einen signifikanten Zuwachs. Eingangs-Terminals/Kameras detektieren Personen, welche keine Maske tragen und schlagen entsprechend Alarm oder verhindern die gewünschte Türöffnung zu einem geschützten Bereich.

Ähnliche Anwendungen gibt es u. a. auf Baustellen um z. B. das Tragen von Bauhelmen und/oder Schutzkleidung zu kontrollieren. Erkennt die KI das Fehlen von Helmen/Schutzkleidung kann ein Alarm ausgelöst werden.

Personen- und Objekterkennungsverfolgung
KI-Algorithmen können Personen in Videostreams erkennen und verfolgen. Dies ist besonders nützlich in Menschenmengen oder bei der Verfolgung verdächtiger Personen auch über mehrere Kameras. KI kann Fahrzeuge anhand von Form, Farbe und anderen Merkmalen identifizieren und verfolgen, was besonders bei der Verkehrsüberwachung und Fahndung nach gestohlenen Fahrzeugen nützlich ist.

Hauttemperatur-Analyse
Typische Anwendungen sind auf Flughäfen, öffentlichen Einrichtungen, Shopping Centern, Firmengeländen etc. bei denen Wärmebild-/ oder Thermografie Kameras Personen mit Fieber (erhöhter Temperatur als 37 Grad C) erkennen. Personen mit Fieber lösen hierbei einen Alarm aus, damit Einsatzkräfte reagieren können, oder der Zutritt in geschützte Bereiche verweigert wird.

Personenzählung
Typische Anwendungsfälle ist das Zählen von Personen zum Beispiel im Einzelhandel. Im Wesentlichen geht es darum numerische Daten zur weiteren Auswertung zu erhalten und zumeist wird daher darauf verzichtet auch die Videostreams abzuspeichern.
Mit den DSGVO neutralisierten Metadaten lassen sich diverse weitergehende Applikationen realisieren. Beispielsweise kann die Analyse von physischem Abstandhalten (physical distancing) zwischen Personen umgesetzt werden um Covid-19 Anforderungen zu verwirklichen. Ebenfalls können die Daten verwendet werden um Gruppenbildungen /Personen-Ansammlungen (Crowd Detection) von Personen zu erkennen.

Demographie
Hierbei lassen sich Personen wie Mann, Frau, Junge, Mädchen und in einem bestimmten Bereich auch das Alter von Personen differenzieren und klassifizieren.

Verhaltensanalyse
KI-Algorithmen können verdächtiges Verhalten automatisch erkennen, z. B. das Zurücklassen von Gepäckstücken, unerlaubtes Betreten bestimmter geschützter Bereiche oder ungewöhnliche Bewegungsmuster.

Fortschrittliche KI-Systeme können Gesichtsausdrücke und emotionale Zustände erkennen, was bei Einschätzungen von Bedrohungen und der Analyse von Verhaltensmustern hilfreich sein kann.

Voraussage von möglichen Diebstählen
Einige der Applikationen verwenden KI in der Videoanalyse um nicht erst bei bereits vollzogenen Straftatbeständen zu reagieren, sondern um aufgrund der Verhaltensanalyse von Personen mögliche Straftaten vorherzusagen, noch bevor sie begangen werden.

Eine der Lösungen gleicht zum Training der Algorithmen auffälliges Verhalten von Personen echter Straftatbestände ab, um dadurch mögliche kriminelle Handlungen vorherzusagen. Der Algorithmus wurde hierzu mit Überwachungsdaten von 100.000 Stunden gefüttert, um ihn so zu trainieren, dass er die Mimik, die Bewegungen und die Kleidung der Käufer überwacht und erkennt.

Die erwähnten Beispiele werden oft verwendet, um Echtzeit-Daten von Besuchern in Dashboards darzustellen und können zudem in der Forensischen Suche und zur Auswertung in Business Intelligence Applikationen wichtige Hinweise auf Besucher- oder Kundenverhalten geben, welche zur weiteren Optimierung von Shops, Notausgängen etc. dienen.

Audio Analytics

Nebst KI Lösungen in der Video Security wird die KI zunehmend auch zur Analyse charakteristischer Geräusche verwendet. Eine der Anwendungen ist beispielsweise die Detektion von Aggressionen durch auffälliges Verhalten, wie schreiende Personen in JVA Zellen, Schalterhallen von Behörden usw.
Ein weiteres Anwendungsfeld ist die Glasbruch Detektion in Museen oder Villen oder auch die Detektion von Schüssen im urbanen Umfeld. Wenn die Audio Analytics auslöst, soll dann eine für diesen Ort zuständige Kamera mit Livebildern aufgeschaltet werden.

Verkehrslösungen

Unter Verkehrslösungen wird in diesem Papier im Wesentlichen auf die Verkehrsdatenanalyse Bezug genommen. Gesammelte Metadaten in der Verkehrsdatenanalyse helfen Gemeinden, der Polizei und weiteren öffentlichen Institutionen den Verkehr in ihrem Gebiet zu analysieren und aufgrund der daraus resultierenden Erkenntnisse ebenfalls zu steuern.

Typische Fallbeispiele von Verkehrsdatenanalyse bzw. Einsatzziele für solche Verkehrslösungen basierend auf Video sind:

- Zählung und Kategorisierung der Fahrzeuge
- Verkehrsstromzählung (Quelle/Ziel des Verkehrs)
- Kontrollschildererkennung und/oder Fahrverbotskontrollen/Zufahrtskontrollen
- Verkehrsdichte und Distanz zwischen Fahrzeugen
- Falsch-Parker und Parkplatz-Management
- (Durchschnitts-/Abschnitts-) Geschwindigkeitskontrolle
- Vorhersage des Verkehrsflusses
- Adaptive Ampelsteuerung

KI-Algorithmen als Hilfsmittel bei der forensischen Auswertung von großen Datenmengen

Bei der Langzeitaufzeichnung von enormen Datenmengen kann die KI helfen das Suchen und Finden von bestimmten Ereignissen, Objekten oder Personen zu vereinfachen. Anstatt stunden- oder tagelang Videomaterial manuell durchzusehen, nutzt das System die künstliche Intelligenz, um visuelle Details automatisch zu analysieren und zu verschlagworten. (z.B. Kleidungsfarben, Accessoires, Objekte, etc.).
Durch eine einfache Beschreibung der Suchkriterien – etwa eine Person mit Brille und blauer Hose – kann das System in Sekundenschnelle entsprechende Videosequenzen finden und anzeigen.

Bei der Aufzeichnung großer Videodatenmengen wird jedes Bild durch einen KI-Algorithmus analysiert und mit verschiedenen Attributen versehen. Diese Attribute werden zusammen mit den Bildern als Metadaten mit Zeitstempeln gespeichert, z. B. Geschlecht, Kleidungsfarbe, Brille, Tasche oder weitere Merkmale. Durch die Kombination mehrerer Attribute können Suchergebnisse präzisiert werden.

<u>Ereignisdetektion (ED)</u>

Ein Ereignis-Detektionssystem (ED-System) dient dazu, alle relevanten Ereignisse in einem Tunnel oder auf der offenen Strecke sofort zu erkennen. Sobald ein Ereignis erfasst wird, übermittelt das System automatisch Alarme an verbundene und übergeordnete Systeme. Auf diese Weise können sicherheitsrelevante Maßnahmen rechtzeitig eingeleitet werden – zum Beispiel:

- die Umschaltung der Signalisation,
- die Regulierung oder Abschaltung der Tunnelbelüftung,
- oder die Sperrung einzelner Verkehrsspuren.

Typische Detektionen in diesem Sicherheitsbereich sind u. a.:

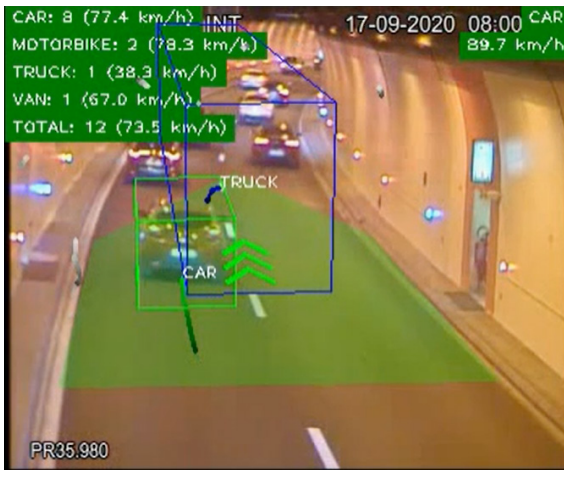
- Stehendes Fahrzeug (stop)
- Langsam fahrendes Fahrzeug (slow)
- Staudetektion (queue)
- Detektion Falschfahrer (wrongway)
- Rauchdetektion oder Bildstörung (smoke/loss of visibility)
- Personen auf Fahrbahn (pedestrian)
- Verlorene Ladung (lost cargo)

Der Einsatz von Systemen mit KI konnte die Genauigkeit der Detektionen im Verkehrsbereich inzwischen signifikant steigern, da zunehmend präziser zwischen Objekten und deren Unterkategorien unterschieden werden kann.

Die nachfolgenden Abbildungen zeigen am Beispiel die Entwicklung der Detektionstechnologie. Sie verdeutlichen, welche Verbesserungen durch die fortschreitende Evolution dieser Technik möglich wurden. Durch den Einsatz von Deep-Learning-Algorithmen kann die Technologie mithilfe künstlicher Intelligenz sowohl Fußgänger als auch Fahrzeuge zuverlässig erkennen und unterscheiden. Darüber hinaus ist sogar eine genaue Kategorisierung der Objekte möglich. Dies führt insgesamt zu einer deutlich reduzierten Zahl an Fehlalarmen.

Vor dem Einsatz von KI (bis ca. 2010) konnte die Verkehrsanalyse technisch bedingt nur Pixel-Änderungen detektieren. Derartige Systeme konnten aufgrund der zweidimensionalen Darstellung typischerweise noch keine Objekte differenzieren.
Erste Systeme setzten bereits Algorithmen in einer „intelligenteren" Kamera ein, ermöglichten aber noch eine ziemlich unpräzise Fahrzeugs-Kategorisierung, welche noch relativ viele Fehlauslösungen generierte, aufgrund von Störgrößen wie z.B. Reflexionen an Wänden (siehe Bild links), Wasserspiegelungen und weitere externe Licht- und Wetterbedingungen.

In einem nächsten Schritt ab 2010 bis und mit 2020 wurden das System perspektivisch auf eine ermittelte 3D-Detektion, was die Genauigkeit der Detektion markant vorantrieb.



Durch das „Object Tracking" (Bild rechts) ist dieser Algorithmus in der Lage, Objekte im Sichtfeld der Kamera zu verfolgen und deren Route zu rekonstruieren. Objekte werden so lange verfolgt, wie sie sich im Sichtfeld der Kamera befinden. Das System verwendet alle Bilder auf der Fahrbahn des Fahrzeugs, um dessen Verhalten zu analysieren.

Der Algorithmus konnte dabei bereits gewisse Fahrzeugs-Klassen PKW/LKW unterscheiden, wobei auch die Fahrzeugs-Klassifizierung nicht zu 100% fehlerfrei funktionierte. Als typische Fehl-Klassifizierung wurden beispielsweise mehrere hintereinander-fahrende Autos als LKW interpretiert (sichtbar als falsch erkannnte

blaue Box „Truck“ im Bild oben rechts). Zudem wurden PKWs nachts oder bei Gegenverkehr häufig fehlinterpretiert und ebenfalls als LKW klassifiziert.

Ab 2020 gelang es dank dem Einsatz weiterentwickelter KI Algorithmen sowohl Fahrzeuge wesentlich zuverlässiger gleichzeitig zu unterscheiden und die Anfälligkeit von Fehlinterpretationen markant zu reduzieren.

Erkannte Objekte werden dabei durch präzisere Tracking-Boxen abgegrenzt (Bild links). Um die Algorithmen zu trainieren, werden große Bilddatensätze aus über 10.000 installierten und im Einsatz befindlichen Videokanälen verwendet.

Um auch unter sehr kritischen und anspruchsvollen Installationsbedingungen noch bessere und akkuratere Ergebnisse zu erhalten, werden die Algorithmen in den Labors des Herstellers laufend weiter trainiert, um z. B. Wetter- sowie saisonale Einflüsse zu neutralisieren.

Durch den Einsatz von Künstlicher Intelligenz (KI) und Deep Learning können moderne Systeme Fahrzeuge und Personen mit bislang unerreichter Genauigkeit unterscheiden. Zudem sind sie in der Lage, typische Störfaktoren wie Nebel, Schatten, Wasserflächen oder Reflexionen zuverlässig auszublenden. Dadurch entstehen deutlich weniger Falschauslösungen – sowohl im Tunnel als auch auf freier Strecke.

Die Erfahrungen von Verkehrsanalyse-Systemen mehrerer Hersteller zeigen, dass eine bis um den Faktor 10 geringere Falschalarmrate, bei signifikant reduziertem Aufwand an Konfiguration und nachträglichen Optimierungsphasen im Betrieb durchaus erreichbar ist. Diese sehr positive Entwicklung ist momentan klar im Bereich der Personen- und Transportmittel zu erkennen, wo KI sehr effektiv die Detektion präzisiert und unterstützt.

Neben Personen und Fahrzeugen spielt im Verkehrsbereich auch die Erkennung von Objekten eine wichtige Rolle – etwa bei der Detektion von „verlorener Ladung“ oder von Gegenständen auf der Fahrbahn, die ein Hindernis oder ein Sicherheitsrisiko für Verkehrsteilnehmer darstellen können. Darüber hinaus ist auch die Erkennung von Sichtbeeinträchtigungen, beispielsweise durch Rauch, von Bedeutung. Diese können durch Reinigungsarbeiten, Brände, Nebel oder Verschmutzungen entstehen und stellen insbesondere in Tunneln einen wichtigen Sicherheitsaspekt dar.

Im Gegensatz zur Erkennung von Personen oder Fahrzeugen kommt bei der Detektion von verlorener Ladung, Objekten oder Rauch nicht das Modul „Neuronales Netzwerk“ zum Einsatz, sondern das Modul „Änderungsdetektor“. Dieses zusätzliche Modul erstellt ein Modell der Szene – den sogenannten Hintergrund – und aktualisiert ihn kontinuierlich. Durch den Vergleich dieses Modells mit den jeweils neuen Kamerabildern kann der Änderungsdetektor ungewöhnliche Situationen erkennen, zum Beispiel das Auftreten von Rauch oder das Vorhandensein von Objekten auf der Fahrbahn.

7. Voraussetzungen für einen erfolgreichen Einsatz der KI in der Videosicherheitstechnik

Kamerastandort und situative Rahmenbedingungen
Die besten KI-basierten Produkte werden immer dann an ihre Grenze stoßen, wenn die eingesetzte Kamera/Objektiveinheit nicht optimal für die geplante Anwendung ausgesucht wurde. Hier sind insbesondere die gewählten Kamerastandorte aber auch Auflösung und Objektivgüte, Betrachtungsabstand, Blickwinkel, Beleuchtung und Neigungswinkel der Kameraeinheit zu erwähnen. Fehler, die bereits hier in der Planung und Ausführung gemacht werden, können auch von der besten KI-Technologie nicht kompensiert werden. Darüber hinaus ist es sinnvoll den Hinweisen/Empfehlungen der Hersteller zur Planung und Umsetzung zu folgen.

Vertrauen in die Lieferkette der KI-Lösung
Weil die Versprechen der Industrie auf eine vollautomatisch-korrekte Erkennung und Detektion von „vorab als wichtig definierten Ereignissen" basieren, ist Vertrauen des Betreibers in die Lieferkette der angebotenen Lösung unabdingbar.

Dieses Vertrauen kann durch gemeinsame Gespräche vorab zwischen Betreiber, Errichter und dem Hersteller und durch das Abfragen bestimmter Parameter aufgebaut werden:

- Woher stammen die Trainingsdaten?
- Wer hat trainiert?
- Wie lange wurde trainiert?
- Auf welchen Datenmengen basierte das Training?
- Können verbesserte Algorithmen nachträglich in das System aufgespielt werden?
- Wurden die Algorithmen unter allen Bedingungen trainiert?
  - Sommer/Winter
  - Tag/Nacht
  - Gutes Wetter/Schlechtes Wetter
- Was passiert, wenn das Ergebnis weit von den Erwartungen abweicht?
- Kann der Algorithmus unter strenger Einhaltung der DSGVO durch das Einspielen von Echtdaten aus der eigenen Applikation verbessert werden?
- Kann der Hersteller/Entwickler zusagen, dass seine Produkte/Lösungen frei sind von „Adversarial Attacks" (Angriffe während der Trainingsphase, um direkt veränderte Trainingsdaten zwecks Korrumpierung oder negativer Beeinflussung des Algorithmus einzubauen)?

Wichtige weitere Hinweise zur Gesprächsvorbereitung
In der Videosicherheit wird häufig mit Begriffen wie „intelligente Kamera" oder „intelligenter Algorithmus" gearbeitet. In diesem Zusammenhang muss aber darauf hingewiesen werden, dass es aktuell weder klare Prüfstandards noch Normen gibt, die diese Begriffe exakt definieren.

I. d. R. wird die eingesetzte KI in der Videosicherheit in der genutzten Anwendung nicht automatisch weiterlernen. Das bedeutet, dass mögliche Verbesserungen der Auswertung mittels Soft-/Firmware-Update auf die genutzte Plattform aufzuspielen sind. Hier ist es empfehlenswert mit dem Anbieter vorab abzuklären, unter welchen Umständen/Bedingungen diese Soft-/Firmware-Updates zur Verfügung gestellt werden.

8. Detektionsgenauigkeiten

Viele Anbieter von KI-basierten Lösungen machen Aussagen über die Genauigkeit der Detektion. Was aber bedeutet eine Detektionsgenauigkeit von 95% oder mehr?

In der folgenden Tabelle wird die Anzahl der Fehler* visualisiert, damit der interessierte Leser sich ein besseres Bild über die Aussagen der verschiedenen Anbieter machen kann:

	Anzahl Detektionen		
Detektionsgenauigkeit	**1.000**	**50.000**	**100.000**
95%	50 Fehler	2.500 Fehler	5.000 Fehler
99%	10 Fehler	500 Fehler	1.000 Fehler
99,9%	1 Fehler	50 Fehler	100 Fehler

Die Interpretation dieser Tabelle hängt selbstverständlich von der Applikation ab und der Tatsache, ob bspw. die Detektion eine Aufzeichnung triggert, oder aber ob die Anzahl an dargestellten Fehlern als Alarmsequenzen an einen Zentralisten in einer Notruf- und Serviceleitstelle (NSL) zwecks Alarmbearbeitung gesendet werden.

***Definition Fehler in diesem Kontext sind:**

- Falschalarm (FP – False Positive): die Videoanalyse löst einen Alarm aus, obwohl kein definiertes Ereignis vorliegt.
- Ausbleibender Alarm oder negativer Falschalarm (FN – False Negative): die Videoanalyse löst keinen Alarm aus, obwohl ein definiertes Ereignis vorliegt.

2 Beispiele zur Veranschaulichung der Fehlerquote vs. Anwendung:

- Wenn ein KI-basiertes Gesichtserkennungssystem zur Identifikation einer kriminellen Person genutzt werden soll, würde es bei einer Genauigkeit von 99,9% bei 100.000 Gesichtern immer noch eine „False Positive"-Rate von 100 falschen Personen geben. Das ist relativ viel, wenn dies 100 falsche Personenzugriffe zur Folge hätte!
- Wenn ein KI-basiertes Verkehrsdatenanalysesystem zur Identifikation der Anzahl von Fahrzeugen genutzt werden soll, würde bei einer Genauigkeit von 99,9% bei 100.000 Fahrzeugen eine Fehlerquote von 100 falsch oder nicht erkannten Fahrzeugen herauskommen. Das ist relativ wenig, weil für diese Zählung eine Fehlertoleranz von 0,01% durchaus akzeptabel wäre!

Entscheidende Fragen in diesem Zusammenhang sind:

- Welche Fehlerrate ist in der jeweils geplanten Anwendung akzeptabel?
- Wie kam der Hersteller / Entwickler auf die angegebene Detektionsgenauigkeit?
- Unter welchen Bedingungen kamen die angegebenen Detektionsgenauigkeiten zustande?
  - Projektierungshinweisen / Empfehlungen der Hersteller beachten
- Wie zuverlässig sind die Angaben der Hersteller / Entwickler tatsächlich?
  - Gelten diese Angaben auch für meine Anwendung?

9. Handlungsempfehlungen

Zum Thema KI in der Videosicherheitstechnik fehlen noch Industrie- oder Normenstandards als auch ethische Standards. Dies ist ein Grund, warum viele Begriffe noch nicht eindeutig definiert wurden und es deshalb – abhängig vom Hersteller/Entwickler – unterschiedlich interpretierbare Aussagen und Begrifflichkeiten gibt. Es muss daher davon ausgegangen werden, dass es bei der Weiterentwicklung und in der Nutzung der KI basierten Analyse in der Videosicherheitstechnik noch einige Lernprozesse geben wird.

Allgemeine Textphrasen in marketingtechnisch aufwändig gestalteten Broschüren sollten eine Kaufentscheidung nicht alleine beeinflussen. Aufgrund der Komplexität der gesamten Thematik wird empfohlen, dass Anbieter dieser Technologie nach Möglichkeit eine Teststellung im Objekt anbieten (im Sinne eines "Proof of Concept" (PoC)), um mit Echtdaten zu verifizieren, ob die Erwartungen des Betreibers durch die Technik des Herstellers erfüllt werden können. Dabei ist darauf zu achten, dass exakt die Kamera- und Beleuchtungstechniken als auch alle weiteren Rahmenbedingungen für einen optimalen Einsatz aller Gerätschaften, die später zum Einsatz kommen sollen, in der Testphase aufgebaut werden.

Sollen mehrere KI-basierte Lösungen gleichzeitig getestet werden, ist zwingend darauf zu achten, dass die gleichen Kamerasignale zur Speisung in die KI genutzt werden, um einen echten Vergleich machen zu können. Idealerweise sollte die Testphase über mehrere Tage/Wochen laufen, um unter verschiedenen Bedingungen wie z.B. Tag/Nacht und unter unterschiedlichen Wetterszenarien ein entsprechendes Ergebnis liefern zu können.

10. Ausblick

In der Zwischenzeit bieten fast alle Marktteilnehmer KI basierte Lösungen für ihre VSS-Produkte an und es wird zukünftig noch sehr interessante Entwicklungen auf diesem Gebiet geben. Mittel- bis langfristig wird KI helfen die Anzahl der Falschalarme deutlich zu reduzieren, was insbesondere bei Fernaufschaltungen von Videosystemen auf 24/7 besetzte Notruf- und Serviceleitstellen (NSL) zu einer wesentlichen Entlastung der Zentralisten führen wird, da diese sich besser auf die echten Alarme konzentrieren können. Dies wird natürlich dazu führen, dass zukünftig die Anzahl der Videoaufschaltungen erhöht werden kann, ohne dass Personal in der NSL aufgestockt werden muss. Dies führt zu einer Kostenreduzierung auf allen Ebenen, so dass auch von einer höheren Marktakzeptanz von Videoaufschaltungen zukünftig auszugehen ist.

Selbst wenn die VSS nicht auf eine NSL aufgeschaltet werden soll, sondern die KI-Kamera-Umschaltungen oder ereignisgesteuerte Aufzeichnungen triggern soll, wird es aufgrund der Falschalarmreduzierung zu enormen Zeitersparnissen bei forensischen Auswertungen kommen.

Bildverzerrungen aufgrund von extremen Weitwinkelobjektiven werden mittels KI basierter Bildkorrektur zukünftig der Vergangenheit angehören; Umwelteinflüsse (z.B. Regen, Wind, Pfützen, Gegenlicht, etc.) werden nicht mehr zu Falschalarmen führen; selbst Insekten direkt vor der Kameralinse werden nicht mehr zwingend zu falschen Alarmen führen.

Es werden größere Sprünge in der Verarbeitungsgeschwindigkeit und in der Erkennungsgenauigkeit zu erwarten sein. Aufgrund der Tatsache, dass sich die KI in der Videosicherheitstechnik noch in einer relativ frühen Phase befindet, muss man allerdings zum jetzigen Zeitpunkt davon ausgehen, dass diese Technik den Menschen assistieren, aber nicht ersetzen kann. Es ist daher anzunehmen, dass noch einige Jahre vergehen werden, bis völlig autark - und korrekt agierende, KI-basierte Videosicherheitssysteme erhältlich sein werden.

11. Quellen und weiterführende Informationen

Der BHE bietet umfangreiche Informationen zum Planen, Errichten und Betreiben von Videoanlagen, beispielsweise den „Praxisratgeber Videosicherheit“ oder vielfältige Informationspapiere. Wissenswertes zum Thema Video lernen Errichter, Planer und Betreiber in den verschiedenen BHE-Seminaren.

[1] Dallmeier 2019 „Videotechnik und KI: Vier Thesen von Dallmeier“, *https://www.sicherheit.info/videotechnik-und-ki-vier-thesen-von-dallmeier*

[2] *https://www.ibm.com/think/topics/machine-learning*

[3] *https://de.wikipedia.org/wiki/Deep_Learning*

[4] *https://www.axis.com/de-de/learning/web-articles/video-analytics*

[5] *https://www.ionos.de/digitalguide/server/knowhow/was-ist-on-premises/*

[6] *https://www.enterpriseitworld.com/computing-for-video-analytics-cloud-on-premise-or-hybrid/*

[7] *https://www.computerwoche.de/a/artificial-intelligence-das-training-macht-den-unterschied,3546899*

BHE Bundesverband Sicherheitstechnik e.V. **Feldstr. 28 66904 Brücken** **Telefon: 06386 9214-0 E-Mail: info@bhe.de** **Internet: www.bhe.de**