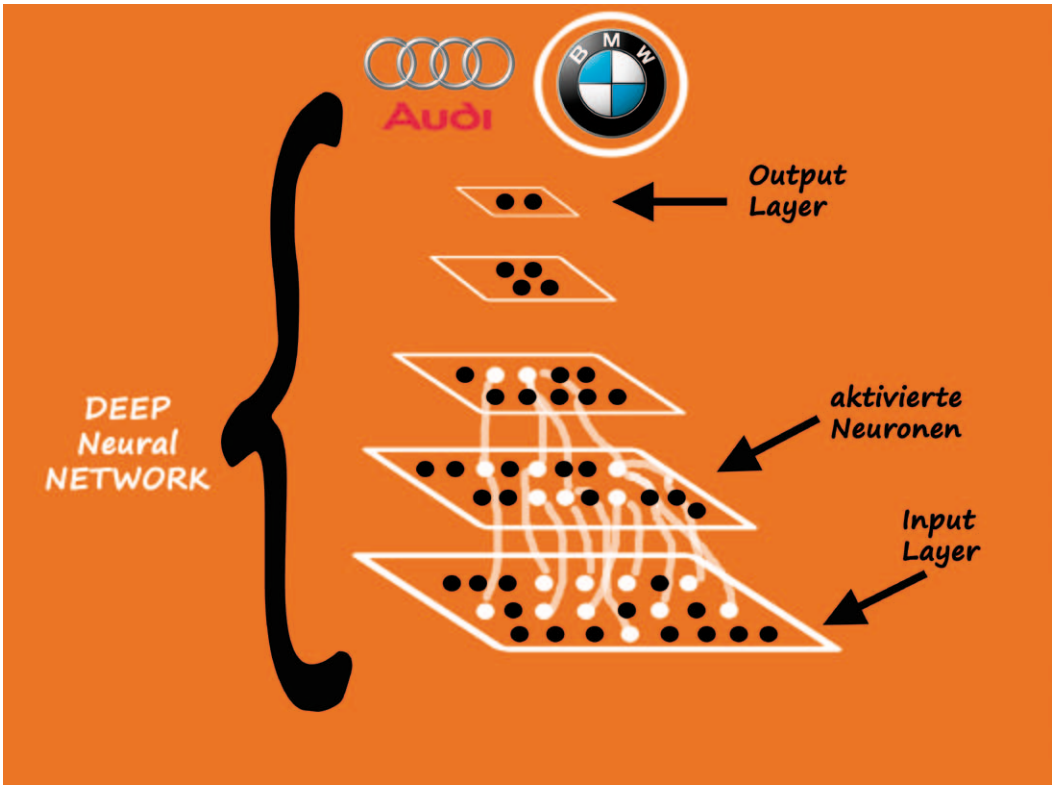
Deep Learning für die Qualitätskontrolle



Unser Verstand lässt visuelle Erfassung sehr einfach aussehen. Ein Mensch kann mühelos einen Panda von einem Eisbären unterscheiden, kann ein Schild lesen oder das Gesicht eines anderen Menschen erkennen oder Obst sortieren. Aber diese Aufgaben sind ziemlich schwierig mit einem Computer zu lösen. Es scheint nur für den Menschen so einfach, weil unser Verstand so exzellent darin ist Bilder zu verstehen.

In den letzten Jahren hat Machine oder Deep Learning stark aufgeholt und ist mit neuen Algorithmen und verbesserter Rechenleistung diese Schwierigkeiten angegangen. Besonders durch das Modell Convolutional Neural Network (CNN), das vor allem in der industriellen Bildverarbeitung von großer Bedeutung ist, sind nun auch verwertbare Ergebnisse für schwierige visuelle Erkennungsaufgaben erhältlich, die nicht nur den menschlichen Fähigkeiten ähnlich sind, sondern diese auf manchen Gebieten auch übertreffen. Ein Beispiel dafür ist Make and Model.

Make and Model – Autos erkennen besser als der Mensch?

Hier geht es um die Erkennung von Automarke und -typ mit Hilfe industrieller Bildverarbeitung. „Make bezieht sich auf den Hersteller eines Fahrzeugs – den „Maker“, wie z.B.: Ford, Toyota, Honda usw. „Model“ ist der spezifische Name, den der Hersteller dem Auto gegeben hat, wie z. B. Focus, Accord, Pathfinder, etc. Für Menschen ist das eine sehr einfache Aufgabe, besonders für Autoliebhaber. Ein Mensch kann Autos an gewissen Hauptmerkmalen wie z. B. Logos, Kühlerfiguren, Form oder Aufschrift, usw. erkennen. Für Computer war dies bisher aufgrund der visuellen Komplexität eine schwierige Aufgabe.

Was leistet Make and Model?

Es kann nicht „einfach nur“ Automarken erkennen und zwischen Audi und VW unterscheiden, es kann auch das Baujahr des selben Modells erkennen, etwas das für den Menschen auch sehr schwierig ist. Allerdings sehen nicht nur Autos der selben Marke ähnlich aus, sondern auch ähnliche Typen anderer Hersteller, beispielsweise die vielen Modelle der SUVs. Auch hier ist eine eindeutige Klassifizierung seitens des Systems möglich. Make-and-Model erkennt also nicht nur das Offensichtliche – wie Autos wie z. B. a Bugatti Veyron „Hypercar“, mit einem sehr speziellen Aussehen – sondern Unterschiede, die selbst für das menschliche Auge schwer zu erkennen sind. Fließhecks die z. B. bei Marken wie Nissan und Toyota sehr ähnlich aussehen, können immer noch unterschieden werden und das auch bei Bildaufnahmen von einer Straße mit vorbeifahrenden Autos, wie Autobahnen oder anderen vielbefahrenen Straßen.

Die Erkennung ist auch möglich, wenn das System ein wesentliches Merkmal, beispielsweise das Logo nicht erkennen kann. Die Auswertung vieler eingelernter Merkmale

Kurz gefasst

Deep Learning gibt es schon länger, aber in den letzten Jahren ist es so leistungsfähig geworden, dass es bestimmte Aufgabenstellungen besser und fehlerfreier als der Mensch lösen kann.

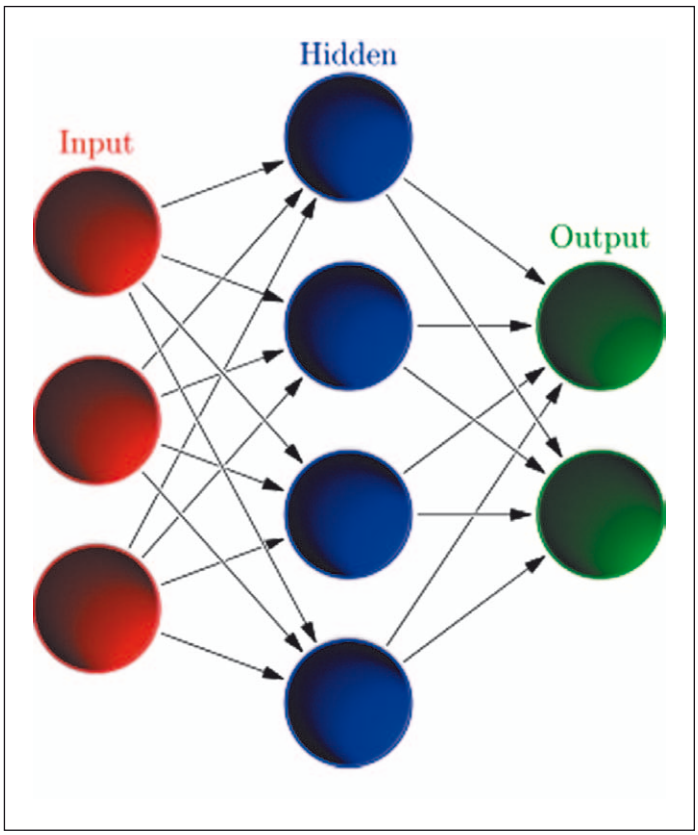


Bild 1: Neurales Netzwerk: Layer L1 = Input Layer, L2/L3 = Hidden Layer, L4 = Output Layer

macht es möglich, dass auf einzelne markante Details verzichtet werden kann, ohne das dies zu einem fehlerhaften Ergebnis führt. Bei der Identifikation des Baujahrs unterschiedlicher Modelle kommt der Mensch meist an seine Grenzen, auch Autoliebhaber. Wir können nicht mehr jedes Baujahr und Modell von jedem in Europa oder Asien oder USA fahrenden Auto identifizieren. Hier ist das System im Vorteil.

Wie kann Machine/ Deep Learning diese Unterschiede erkennen?

Die größte Herausforderung ist es eine feinkörnige Klassifizierung zu erreichen und der Grund dafür sind unumstritten die feinen Unterschiede zwischen den unterschiedlichen Klassen. Um diese winzigen Unterschiede zu lernen ist ein sehr großer Datensatz notwendig. Man sollte bedenken, dass Deep Learning unter Bedingungen wie wenig Zeit, Rechenleistung, oder Daten, nicht durchzuführen ist. Machine Learning ist Lernen von Beispielen und Erfahrung welchen Algorithmen zugrunde liegen. Beispielsweise ist Transfer Learning eine Machine-Learning-Technik, welche die gelernten Klassifizierer für neue Aufgaben umwidmet. Transfer learning ist eine Methode, die sich die Lösung eines Problems merkt und das dadurch erlangte Wissen auf die Lösung verwandter Probleme überträgt. Beispielsweise kann das System nach dem Lernen von Automarken auch LKWs identifizieren.

Für CNNs (Convolutional Neural Networks) wird ein Basisnetzwerk mit einem Basisdatensatz trainiert um Gewichtung und Eigenschaften zu generieren. Am Ende von Transfer Learning entsteht ein Klassifizierer, welcher auf den neuen Datensatz anwendbar ist. Dies bedeutet weniger Aufwand als ein komplettes Netzwerk neu einzulernen. Dieses durchaus leistungsfähige Tool kann ein großes Zielnetzwerk trainieren und gleichzeitig Data-Overfitting (Überanpassung) minimieren. Beim Trainingsprozess steigt die Erkenntnisleistung parallel zur Anzahl eingelernter Bilder. Ist die Sättigungsphase erreicht, nimmt die Erkenntnisleistung wieder ab, weil sich die Datenrepräsentation des Rechners zu sehr an die Trainingsdaten anpasst und nicht mehr an den zugrundeliegenden Formen orientiert. Dieser Prozess wird als Überanpassung bezeichnet. Selbstverständlich sind auch andere Techniken zum Trainieren anwendbar.

Was sind Neural Networks oder auch Convolutional Neural Networks (CNN)?

Neurale Netze sind die Basis für Machine/ Deep Learning mit mehreren Layern. Am besten stellt man sich eine Familie an Modellen vor, die sehr „loose" vom menschlichen Gehirn inspiriert wurden, und Funktionen näherkommen die von einer großen Anzahl an Inputs abhängen. Bei Make and Model gilt je größer der Datensatz an klassifizierten Bildern ist desto besser.

Ein Neurales Netz sind Beispiele von nichtlinearen Hypothesen, wo das Modell noch viel komplexere Beziehungen lernen kann zu klassifizieren. Statt dem Computer einzeln jede Möglichkeit der Automodelle einzulernen, wird das Machine Learning Programm einfach mit mehreren Tonnen an existierenden Bildern gefüttert.

Ein Neuronales Netz wird durch künstliche Neuronen gebildet, die in Layern (Schichten) angeordnet sind. Dabei gibt es drei Arten von Layern:

- Input Layer: sucht nach allen Pixeln in einem Bild
- Hidden Layers: können gewissen Merkmale, Aspekte, Formen, Texturen, etc. erkennen
- Output Layer

Neurale Netze mit mehr als zwei „Hidden Layers" können als Deep Neural Networks bezeichnet werden. Je mehr Layer desto komplexere Muster können erkannt werden. An einem Beispiel von zwei Layern kann man sich die Verbindungen zwischen den Neuronen als Muster vorstellen, welches gelernt wird.

Unterschied zum menschlichen Gehirn

Um ein riesiges Neuronales Netzwerk zu bilden und es Artificial Intelligents (AI) zu nennen, müssen u. a. folgende Punkte erfüllt werden:

- Die künstlichen Neuronen funktionieren ganz anders als die Nervenzellen in einem Gehirn
- Ein menschliches Gehirn hat 100 Milliarden Neuronen und 100 Billionen Verbindungen (Synapsen) und operiert mit ca. 20 Watt (genug um eine schwache Glühbirne zu betreiben) – im Vergleich dazu hat ein Neuronales Netz nur 10 Millionen Neuronen und 1 Mil-

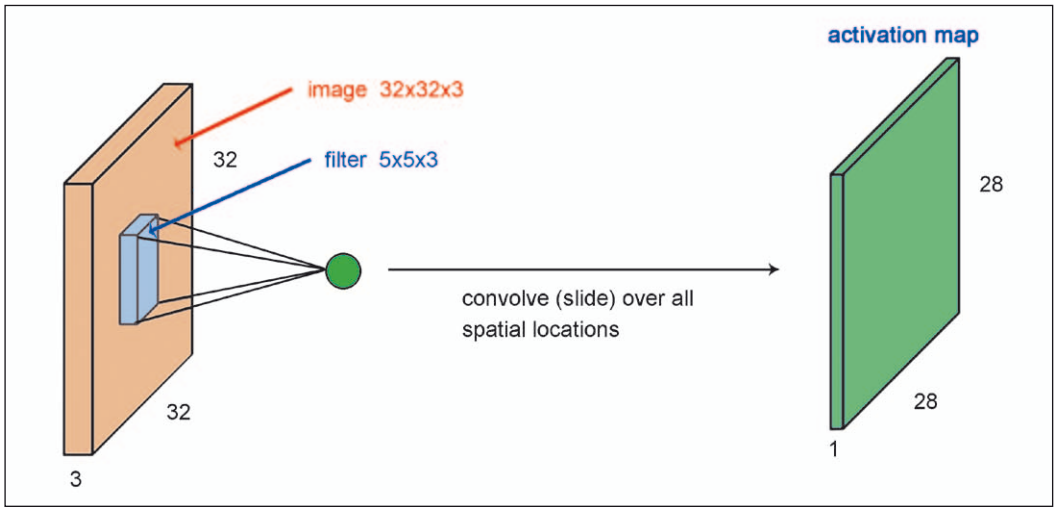


Bild 2: Convolve over all spatial locations

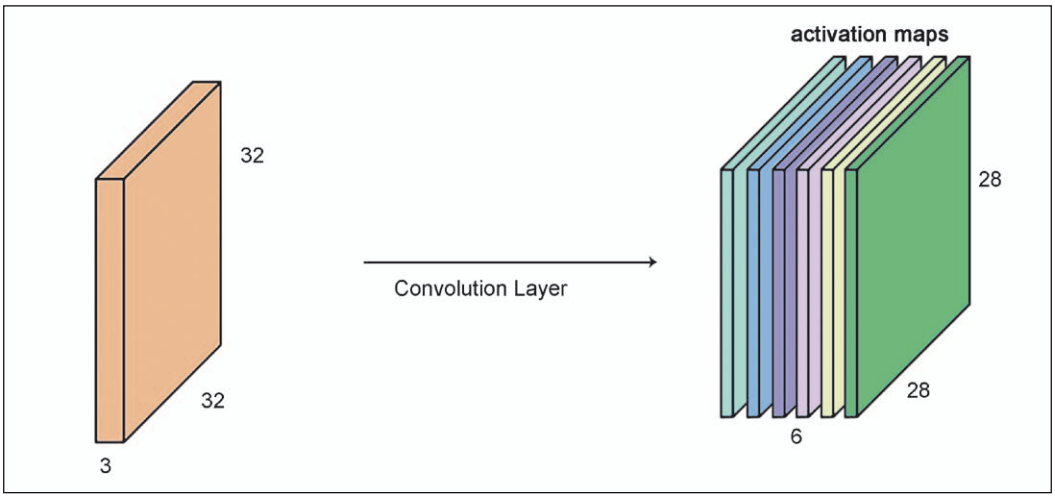


Bild 3: Convolution Layer

liarde Verbindungen auf 16.000 CPUs (ungefähr 3 Millionen Watt)
- Das Gehirn ist beschränkt auf 5 Typen von Input Daten von 5 Sinnen
- Kinder lernen anders als Netzwerke: Sie lernen nicht, was ein Auto ist, dadurch dass sie 100.000 Bilder durchsehen welche mit „Auto“ und „nicht Auto“ gekennzeichnet sind - aber so lernt Machine Learning
- Wahrscheinlich lernen wir nicht dadurch, dass wir die partielle Abweichung eines jeden Neurons in Beziehung zu unserem initialen Konzept errechnen, aber Machine Learning

Convolutional Neural Networks (CNN)

Der Vorteil von CNNs ist, dass sie nicht nur Gewichtung der Features, sondern auch die Features selbst lernen können. Zusätzlich haben CNNs eine hochmoderne Genauigkeit zur generischen Bildklassifizierung erlangt. Die wichtigste Operation an Convolutional Neural Networks sind die Convolution Layers. Stellen wir uns ein 32 x 32 x 3 Bild vor. Wenn wir dieses jetzt mit einem Filter (= Faltmatrix) von 5 x 5 x 3 „convolve“ („falten“), ist das Ergebnis ein Aktivierungsmarker von 28 x 28 x 1. Dabei muss der Filter immer die selbe Tiefe haben wie das Input-Bild. Die Faltmatrix sucht nach einem bestimmten Merkmal auf dem Bild, d. h. es sucht nach einem Muster im gesamten Bild mit nur einem Filter (Bild 2).

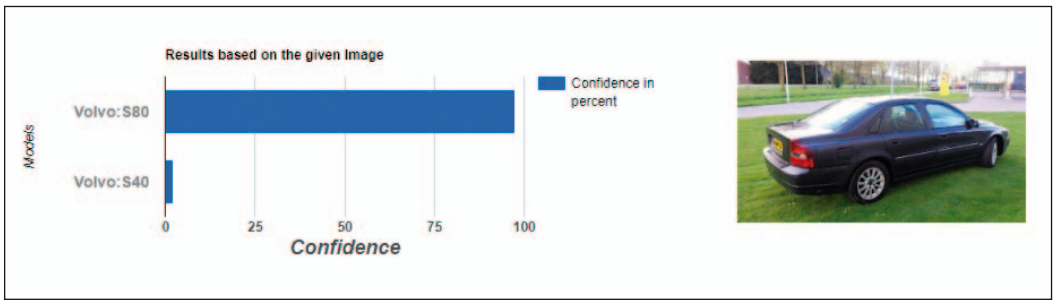
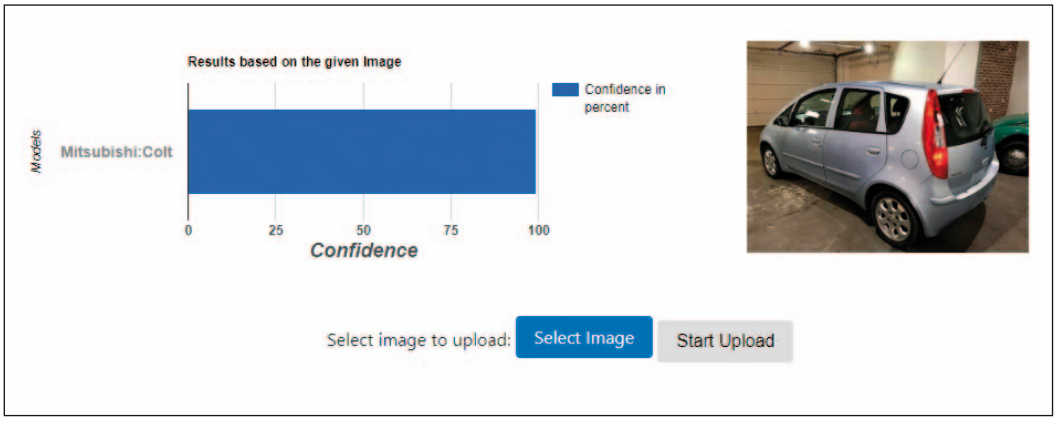
Wenn wir wollen, dass unser Convolution Layer nach sechs verschiedenen Merkmalen sucht, benötigen wir sechs Filter. Jeder Filter sucht nach einem bestimmten Muster auf dem Bild (Bild 3).

Die Convoluiton (Faltung) ist eine lineare Operation. Die Faltung wird solange fortgesetzt, bis das System auf ein nichtlineares Layer trifft. Dies ist normalerweise ein Relu. Ein anderer wichtiger Punkt ist, dass es für die Faltmatrix irrelevant ist wo im Bild sich das Muster befindet. Das Muster wird trotzdem erkannt.

Wo wird Make and Model eingesetzt?

Make-and-Model-Erkennung von Autos ist heutzutage ein gefragtes Element für automatische Vision basierte Systeme, z. B. zur Verkehrsüberwachung, Überwachungskameras zur Strafverfolgung oder Zugangskontrollen für Parkplätze, Gebäude oder auch Sperrgebiete. Zugangskontrollen haben durch das gesteigerte Sicherheitsbewusstsein der letzten Jahre immer mehr an Bedeutung gewonnen. In Überwachungssystemen kann ein Videostreaming dem Sicherheitspersonal helfen, die Fahrzeuge anhand der Marke, dem Modell oder der Farbe zu verifizieren und dieses Fahrzeug dann über die gesamte Videokamerabandbreite verfolgen. Beispielsweise ist es für die Strafverfolgungsbehörden Voraussetzung, dass sie ein Auto in Verbindung mit einem Verbrechen bzw. ein verdächtiges Fahrzeug über mehrere Millionen von Verkehrsaufnahmen verfolgen können, und zwar manchmal nur durch Beschreibungen von Augenzeugen. Wenn man von vornherein Make and Model einsetzt, unterstützt dies das Personal und kann die Überführung des Täters beschleunigen. Außerdem kann MMR (make & model recognition) auch an Grenzübergängen und elektronischen Mautstellen oder bei der Verkehrskontrolle und -überwachung, für Statistiken Verkehrsdichte/-stärke eingesetzt werden.

Einsatzbereiche für Make and Model

- Beobachtung und Ermittlung
- Sicherheitsüberwachung
- Verkehrsdaten sammeln
- Strafverfolgung
- Mautstellen

Autorin:
Klara Steinschneider
EVT Eye Vision Technology
www.evt-web.com