

Titel:

AI meets creativity: recent advancements in visual computing, virtual and augmented reality

Künstliche Intelligenz trifft Kreativität: Die neuesten Entwicklungen im Bereich Visual Computing, Virtual und Augmented Reality

Das Deutsche Forschungszentrum für Künstliche Intelligenz (KI) wurde im Jahr 1988 mit der Zielsetzung, eine Brücke zwischen der rein akademischen und universitären Welt zu der praxisorientierten Wirtschaft und Industrie zu bilden, gegründet. Sowohl Universitäten, an denen DFKI-Standorte angesiedelt sind, als auch zahlreiche namhafte Unternehmen sind als Gesellschafter am DFKI beteiligt und unterstützen das Institut im Sinne eines „Private Public Partnership“. Der Forschungsbereich Augmented Vision (<http://av.dfki.de/>) beschäftigt sich mit drei Forschungsgebieten, nämlich „3D Computer Vision“, „Body Sensor Networks“ und „Virtual and Augmented Reality“ und entwickelt zu diesen Themengebieten neue Verfahren und dedizierte Softwarelösungen.

KI ist als Schlagwort zur Zeit in allen Munde und wird in naher Zukunft eine wesentliche Rolle in unserem täglichen Leben sowie in der Industrie spielen. Wieso ist KI seit Kurzem so präsent und wichtig? Dies liegt im Wesentlichen an der aktuellen Entwicklung, die „digitale Transformation“ genannt wird. Die erste Phase dieser Transformation bestand darin, Daten von Rechnern lesbar zu machen und Maschinen und Geräte miteinander zu verbinden, um einen effizienten Austausch von Informationen zu ermöglichen. Diese erste Phase wird u.a. von den Technologien „Cloud-Computing“ und „mobiles Internet“ geprägt. Die enorme Ansammlung von Daten und deren Verwaltung führten auch zu dem Begriff „Big Data“. Diese erste Phase ist technologisch gesehen abgeschlossen. Nun geht es darum, die gesammelten Daten effizient zu nutzen und zu verwerten, um daraus einen Mehrwert sowie neue Geschäftsmodelle zu schaffen. Dafür ist es notwendig, dass Rechner und Maschinen diese Daten verstehen und richtig interpretieren. Das heißt, dass wir heute die zweite Welle der Digitalen Transformation erleben, und diese wird von den Technologien aus der Künstlichen Intelligenz und des Maschinen Lernens angetrieben.

Was sind die wesentlichen Verfahren von KI heute? Viele KI-Methoden wurden seit über 20 Jahren erforscht und evaluiert, aber die aktuellen Rechnerkapazitäten und die Verfügbarkeit von großen Datenmengen öffnen ganz neue Möglichkeiten. Sogenannte tiefe neuronale Netzwerke oder „Deep Neural Networks“ können erst heute über weite Datensätze effizient trainiert werden, was deren „Performance“ enorm steigert. Ist die Lernphase abgeschlossen, erkennen diese Netzwerke mit einer extrem hohen Genauigkeit die gelernten Signale aus neuen Daten. Sie übertreffen bei Weitem die klassischen Verfahren aus dem Gebiet der Mustererkennung, die auf einer expliziten (nicht durch Lernen) Formulierung der Erkennungsaufgabe beruhen. Der heutige Erfolg von KI ist vor allen Dingen auf das Gebiet des Maschinen Lernens und das sogenannte Deep-Learning zurückzuführen. Wo und wie diese Verfahren am DFKI eingesetzt werden, wird im Folgenden anhand von Beispielanwendungen deutlich gemacht.

Die Applikation „Augmented Things“ stellt eine neue Architektur und Anwendung von Augmented Reality (AR) dar. AR kann in der heutigen vernetzten Welt und des „Internet of Things“ (IoT) als universelle Mensch-Maschinen-Schnittstelle betrachtet werden. Hierfür wird eine neue Systemarchitektur benötigt. Die Grundidee von „Augmented Things“ ist, dass alle notwendigen Informationen für die Durchführung einer AR-Anwendung lokal auf dem Objekt gespeichert werden. Dem Smartphone oder der AR-Brille liegen keine Informationen vor. Das verwendete AR-Gerät verbindet sich zuerst mit dem Objekt, liest und spielt dann die übertragenen Informationen ab. Jede Maschine, die mit diesen Informationen bestückt ist, ist „AR-Ready“ und kann mit dem Benutzer effizient kommunizieren.

Die Aufbereitung des graphischen Inhalts ist aber heutzutage immer noch aufwendig, insbesondere für anspruchsvolle Anwendungen mit detaillierten Informationen, wie beispielsweise die Benutzer-Assistenz für Reparaturanleitungen. Im Forschungsbereich „Augmented Vision“ wurde Methoden zur automatischen Erstellung von AR-Inhalten durch Beobachtung von Vorgängen mit Hilfe einer Kamera entwickelt. Ein Video eines Reparaturablaufes z.B. stellt das Können des Operators dar und beinhaltet implizit alle Informationen über den durchgeführten Vorgang. Hierbei stellt sich die Aufgabenstellung wie die Informationen aus einem Video extrahiert und strukturiert werden können. Die Umsetzung dieses Digitalisierungsschrittes würde es ermöglichen, Wissen und Expertise einer Person (gewissermaßen!) aufzunehmen, zu strukturieren und daraus Inhalte für neue Trainingssimulatoren oder AR-basierte Assistenzsysteme automatisch zu erzeugen. Der entwickelte Ansatz besteht darin, das Video im ersten Schritt in einzelne Kapitel zu unterteilen, sogenannte „Atomic Actions“. Jedes einzelne Kapitel wird dann analysiert und Handgesten werden extrahiert. Anschließend kann ein Benutzer zu einem gegebenen Kapitel weitere nicht sichtbare Informationen, wie z.B. Texte, Identifikationsnummern von Bauteilen, usw. in die Bilder einfügen. Damit ist der gesamte Arbeitsablauf strukturiert und digitalisiert.

Wenn ein neuer Benutzer eine gleiche Maschine über die AR-Brille anschaut, wird diese erkannt und die passenden Workflow-Informationen dargestellt. Der Benutzer sieht eingeblendet und lagerichtig die Hände des Experten, der die ersten Arbeitsschritte demonstriert. Das System vergleicht die durchgeführte Arbeit mit den vorliegenden Handlungen des aufgenommenen Experten und spielt adaptiv die notwendigen Kapitel ab (siehe: <https://www.youtube.com/watch?v=vtUdafED4rw>).

Weitere Technologien der KI sind sogenannte Bild-basierte 3D Rekonstruktionsverfahren. Pixel unterschiedlicher Bilder einer Szene werden untereinander zugeordnet und ohne weitere Vorabinformationen die 3D Geometrie der Umgebung zurückgewonnen. Die aktuelle Arbeit fokussiert die 3D Erfassung mit Hilfe von 360° Bildern. Der nächste Arbeitsschritt besteht darin, dynamische Szenen aus Videos in 3D zu erfassen und zu modellieren. Die ersten Ergebnisse wurden am Beispiel von bewegten Gesichtern erzielt.

Welche Informationen können noch aus Videos extrahiert werden? Deep-learning-Verfahren erlauben es, z.B. Personen und einzelne Gelenke im Bild zu lokalisieren und anschließend daraus einen 3D animierten Skeleton als „3D Strichmännchen“ abzuleiten. Mit dieser Technologie können z.B. menschliche Interaktionen oder/und Bewegungsabläufe genau analysiert und einfach über Smartphone-Videos durchgeführt werden.

Kann aber ein Rechner scheinbar abstraktere Eigenschaften eines Bildes, wie zum Beispiel den persönlichen Stil eines Malers erfassen und auf ein neues Bild übertragen? Hier sind ein paar Ergebnisse von den sogenannten „Style Transfer“ zu sehen. Ein Foto wird zu einem Van Gogh oder Picasso ähnlich aussehendem Gemälde umgewandelt. Umgekehrt ist es möglich, aus Skizzen photo-realistische Bilder zu generieren. Handelt es sich um Kreativität? Der Rechner kann sehr gut ähnliche Ergebnisse aus Daten reproduzieren. Es ist aber extrem schwierig für einen Rechner, aus unbekanntem Daten etwas Sinnvolles, Neues und Nützliches zu generieren. Das heißt, KI kann als Inspirationsquelle dienen oder Schritte eines Kreativeprozesses übernehmen und unterstützen.

Wann wird der Einfluss von KI besonders groß und sichtbar? KI wird zurzeit auf großen Rechnern erprobt. Der nächste Schritt der aktuellen Entwicklung von KI wird darin bestehen, die Berechnung auf kleine energiesparende und dedizierte „KI-Chips“ zu realisieren. Dadurch kann KI lokal auf Maschinen erfolgen. In dem Projekt „Eyes-of-Things“ (<http://av.dfki.de/projects/eyes-of-things/>) wird ein Miniaturkammersystem mit eingebettetem Deep-Learning Chip und drahtloser Kommunikation entwickelt. Neue Software-Apps können auf diese Eyes-Of-Thing Einheit flexibel hochgeladen werden. Dadurch kann diese autonome Sensoreinheit leicht umkonfiguriert werden und für zahlreiche unterschiedliche Aufgaben benutzt werden. Diese reichen von der Erkennung von Zuständen im Bereich Robotik und Industrie 4.0 bis zu Consumer-Geräten oder Spielzeugen, wie einer Puppe, die das spielende Mädchen erkennt und z.B. dann einen Dialog startet.

Zusammengefasst, ist der aktuellen „Hype“ von KI auf drei zusammenfließende Strömungen zurückzuführen: Daten sind in großer Menge vorhanden, Algorithmen aus dem Bereich Deep-Learning werden intensiv erforscht und neue dedizierte Hardwaresysteme für das Training und die Durchführung solcher komplexer neuronaler Netzwerke stehen zur Verfügung. Neue Applikationen und Geräte werden insbesondere durch die Einbettung dieser Verfahren auf kleine Chips erscheinen. Viele Herausforderungen müssen aber noch überwunden werden; insbesondere muss ein besseres Verständnis über das Trainingsverhalten dieser Netzwerke erlangt werden. Weiterhin muss gesichert werden, wann ein Netzwerk ausreichend trainiert worden ist und neue Methoden gefunden werden, die mit kleinen Datensätzen arbeiten können. Letztendlich soll das Verhalten der Netze besser verstanden und auch sicherheits-relevanten Aufgaben, wie sie beispielsweise bei autonomem Fahren vorkommen, gewährleistet werden, dass die Antwort des Netzwerks absolut korrekt ist. Diese Fragestellungen werden uns in den nächsten Jahren beschäftigen!