Video Intelligentie

Snoek, C.G.M.

Publication date
2019
Document Version
Final published version

Video Intelligentie

Video Intelligentie

Rede

uitgesproken bij de aanvaarding van het ambt van
hoogleraar Intelligent Sensory Information Systems
aan de Faculteit der Natuurwetenschappen, Wiskunde en Informatica
van de Universiteit van Amsterdam
op vrijdag 14 december 2018

door

Cees G.M. Snoek

Dit is oratie 604, verschenen in de oratiereeks van de Universiteit van Amsterdam.

Opmaak: JAPES, Amsterdam

Mevrouw de Rector Magnificus,
Mijnheer de Decaan,
Geachte leden van het curatorium,
Zeer geachte collega' s, studenten, vrienden en familie,

Een half miljard jaar geleden vond de biologische Big Bang plaats, de zogenaamde Cambrische explosie. Tot die tijd werd onze planeet slechts bevolkt door primitieve zeediersoorten die zich het beste laten vergelijken met sponsen, kwallen en wormen. Maar in een relatief kort tijdsbestek werd deze beperkte fauna uitgebreid met de stamvaders en -moeders van al het intelligente leven op aarde. Opeens kregen diersoorten klauwen, tentakels en pantsers (zie figuur 1). Het ontstaan van deze plotselinge dierlijke verscheidenheid plaatste Darwin voor een raadsel en is ook vandaag de dag nog een onopgelost mysterie.

Een mogelijke verklaring voor deze bijzondere gebeurtenis, die ik graag wil geloven, is de 'lichtknop theorie' (Parker, 2016). Deze stelt dat het rond die periode veel lichter werd op onze planeet waardoor dieren een nieuw zintuig ontwikkelden: ze kregen ogen en leerden kijken. Als gevolg van deze nieuwe gave werd uiterlijk een kwestie van eten of gegeten worden. Immers, een worm met stekels is minder makkelijk door te slikken dan een worm zonder. Kortom, de mogelijkheid om te kunnen kijken veranderde zowel de verscheidenheid als het gedrag van intelligent leven. Vandaag staan we aan de vooravond van misschien wel een tweede Cambrische explosie, waarbij ook levenloze voorwerpen leren kijken.

Over drie jaar al, zullen er maar liefst vijfenveertig miljard camera's op de aarde zijn[1]. Dit wordt mogelijk gemaakt door twee voortschrijdende technologische ontwikkelingen. Ten eerste worden computerchips met camera's steeds krachtiger, kleiner en goedkoper. Ten tweede wordt draadloze communicatie steeds sneller en kan ze steeds meer data aan. Al deze voorwerpen met camera's worden digitaal met elkaar verbonden via razendsnelle mobiele netwerken in het *Internet of things that video*. Camera's zijn niet langer beperkt tot die in uw mobieltje, uw babyfoon of uw auto. Ik voorspel u, elk voorwerp met plaats voor een camera en de potentie voor een toepassing zal er één krijgen.

Figuur 1 **Door de Cambrische explosie ontstond een enorme dierlijke verscheidenheid die mogelijk werd veroorzaakt door de ontwikkeling van het gezichtsvermogen. We staan aan de vooravond van misschien wel een tweede Cambrische explosie, waarbij ook levenloze voorwerpen leren kijken.**

Een verrassend voorbeeld vind ik de videopil. U slikt de videopil van 11 tot 27 millimeter door en de camera maakt binnen 24 tot 48 uur een video-opname van uw hele maagdarmkanaal, voordat het pijnloos via de ontlasting uw lichaam weer verlaat. U wordt na afloop wel vriendelijk verzocht om de capsule zelf uit de pot te vissen. De behandelend arts bekijkt de opgenomen videobeelden, wel 24 tot 48 uur per patiënt, en stelt aan de hand daarvan de diagnose. Een actueel voorbeeld is het project Camera in Beeld, waarbij burgers hun buitencamera's vrijwillig kunnen registreren bij de politie. In de periode van oktober 2016 tot juni 2018 groeide dit aantal al van 100.000 tot 200.000 camera's[2]. Na een inbraak, beroving of overval in de buurt kunnen agenten en rechercheurs de beelden opvragen en die pluizen ze dan na in de hoop een verdachte te identificeren. Een wetenschappelijk voorbeeld komt uit de mariene biologie, waarbij speciale videocamera's met extra lichtsensoren de grote diepten van onze oceanen afstruinen om nieuwe zeediersoorten voor het eerst zichtbaar te maken (Maxmen, 2018).

Het nadeel van deze en al die andere camera's is dat de enorme hoeveelheid videobeelden die opgenomen wordt nu nog grotendeels handmatig moet worden bekeken door mensenogen (zie figuur 2). Dit is niet alleen foutgevoelig, tijdrovend en een mogelijke inbreuk op uw privacy, maar stelt ons ook voor een groot probleem: de digitale camera's genereren nu al meer video dan we ooit kunnen en willen bekijken. Camera's moeten daarom niet alleen video creëren maar deze beelden ook volautomatisch interpreteren. Er is, kortom, een noodzaak voor video intelligentie.

In mijn oratie wil ik u meenemen op een reis door dit fascinerende nieuwe vakgebied. Ik zal trachten uit te leggen wat video intelligentie is, wat de grote wetenschappelijke vragen zijn en wat we er nu en in de toekomst mee kunnen. Ook zal ik stil staan bij de betekenis van video intelligentie voor maatschappij en onderwijs.

Figuur 2 **De stortvloed aan digitaal videomateriaal wordt nog steeds grotendeels door mensenogen bekeken. Er is een noodzaak voor kunstmatige video intelligentie.**

1 Boom, roos, vis

De eerste artikelen rond video intelligentie dateren van begin jaren negentig van de vorige eeuw. De focus lag toen vooral op videomateriaal van televisie-uitzendingen, zoals journaals, en hoe deze uitzendingen doorzoekbaar te maken met kunstmatige intelligentie methodieken uit de beeld- en spraakherkenning. Een inspirerend systeem uit die pioniersjaren is Informedia van Carnegie Mellon University uit Pittsburgh (Wactlar *et al.*, 1996), waar ik als jonge promovendus een kleine bijdrage aan mocht leveren. Het was de tijd van het aantonen van de potentie van de nieuwe videotechnologie en het definiëren van het grote wetenschappelijke probleem. Dat probleem werd het *semantic gap*, ofwel het slechten van de kloof die bestaat tussen de informatie die een computer kan extraheren uit een digitale video enerzijds en de interpretatie die een mens geeft aan diezelfde data anderzijds. Anders gezegd, we willen automatisch kunnen begrijpen *wat* er *waar* en *wanneer* gebeurt in een videostroom.

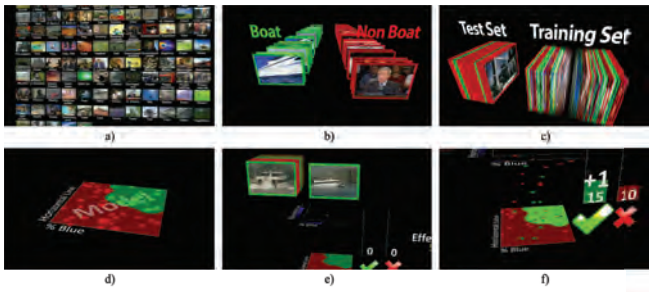
Netflixen op een iPad of televisiescherm wordt in het algemeen als een nogal passieve taak beschouwd. Toch gebruiken onze hersenen al snel 50% van hun capaciteit om videobeelden en geluiden te kunnen interpreteren. Het is daarom niet zo gemakkelijk om dit, kennelijk cognitief complexe, proces te automatiseren. Hoe pak je dit aan? Hoe leer je een computer om video's te lezen? Nou, net zoals we dat doen met onze kinderen, we sturen ze naar de basisschool en na wat spelen in de kleuterklas mogen ze naar groep drie om te leren lezen. Ik leerde destijds eerst het rijtje boom, roos, vis. Zo ging het voor video intelligentie eigenlijk ook.

Alle begin is moeilijk, dus eerst werd voor elk nieuw woord, of ding, een promotie-student ingehuurd die de taak kreeg om een set van regels te verzinnen die voor een computer beschrijven hoe een ding, bijvoorbeeld een boom, er uitziet in video. Een boom heeft een langwerpige stam, bovenop zitten takken, vol met bladeren. De configuratie van die elementen definieerde dan een boom. Ging best goed zolang kerstbomen niet voorkwamen, dan moesten de regels aangepast worden. Ook als na de promotie de hoogleraar nog verder wilde op het onderwerp maar nu interesse had in rozen of vissen, moest weer van voren af aan begonnen worden met het verzinnen en programmeren van nieuwe regels, de oude voldeden immers niet meer. Het moge duidelijk zijn, dat deze manier van werken niet schaalbaar is, je bent simpelweg te lang bezig om vooruitgang te boeken. Bovendien zijn regelgebaseerde methoden, naast hun beperkte schaalbaarheid, nogal broos en beperkt. Maar, al doende leert men.

Waar in de astronomie de grote wetenschappelijke doorbraak kwam door niet langer de aarde maar de zon centraal te stellen, zo werd dat in de video intelligentie bereikt door niet langer het 'ding' maar de videobeelden van het ding centraal te plaatsen. Met de afbeeldingen wordt het mogelijk om de beslisregels te leren uit beelddata in plaats van deze te programmeren uit je hoofd. Dit is een uiterst krachtig paradigma gebleken want het stelt ons in staat om de video intelligentie grotendeels afhankelijk te maken van voorbeelden. Dat wil zeggen: voorbeelden gaan in de machine en voorspellingen of het ding in het beeld zichtbaar is komen er weer uit. Op die manier is in principe elk ding in videobeeld te herkennen, zo lang er voldoende voorbeelden voorhanden zijn.

Hoe werkt dat dan? De standaard aanpak voor het geautomatiseerd herkennen van beelden begint met een bepaald visueel ding bijvoorbeeld een boot. De set van gelabelde video's wordt verdeeld in een training set en een test set. De training set wordt gebruikt voor de optimalisatie van de software en voor het aanleren van een zogenaamd statistisch model dat de visuele weergave van het betreffende ding vastlegt in een wiskundige formule. De test set wordt gebruikt om de mate waarin het model beelden herkent te evalueren door zijn voorspellingen te vergelijken met de oorspronkelijke afbeelding (zie figuur 3).

Figuur 3 **Video intelligentie in een notendop: a) begin met selectie van een visueel ding om te herkennen, in dit geval een boot, b) verzamel positieve en negatieve videofragmenten met en zonder boten, c) verdeel de verzamelde fragmenten in een training en een test set, d) leer op basis van de voorbeelden uit de training set een model dat zo goed mogelijk de positieve van de negatieve voorbeelden kan scheiden, e) voorspel hoe goed het model in staat is om de fragmenten in de (ongeziene) test set correct te herkennen, f) bereken een score om de kwaliteit van het model te meten.**



De machines voor het voorspellen van dingen in videobeelden zijn de afgelopen vijftien jaar ontzettend precies geworden. Een belangrijke motivator voor het voortstuwen van de prestaties is de internationale TRECVID-competitie voor videozoekmachines die het Amerikaanse Instituut voor Standaarden en Technologie startte in 2001[3]. Dat was toevalligerwijs ook het jaar dat ik begon met mijn promotieonderzoek. De organisatoren van de TRECVID-competitie observeerden dat onderzoek naar video intelligentie wijdverspreid was, maar dat elke onderzoeker zijn intelligente video machine zelf evalueerde op eigen videodata. Het was daarom rond 2001 bijzonder moeilijk om machines met elkaar te vergelijken en verschillende methodieken te repliceren. Elke onderzoeker vond zijn eigen werk ook toen al fantastisch.

Om progressie in het onderzoek te promoten stelde de competitie grote video datasets beschikbaar, dat was hard nodig want videodata was aan het begin van deze eeuw helemaal niet makkelijk te verkrijgen. Het is tegenwoordig nog maar moeilijk voor te stellen, maar YouTube kwam echt pas in 2005 en de eerste iPhone dateert van 2007. Ook definieerde de competitie gestandaardiseerde taken, waarbij elke deelnemer hetzelfde probleem moest oplossen op dezelfde data. Resultaten werden ingeleverd bij de organisatie en onafhankelijk geëvalueerd. Maar de beste zet van de organisatoren was om elke deelnemer te verplichten de kaarten na afloop open en bloot op tafel te leggen. Door dit open innovatie model kon van elke deelnemende machine worden achterhaald waarom het werkte of faalde. Op die manier leerden alle deelnemers razendsnel van elkaar. Een 'geleerde' gemeenschap rond video intelligentie was een feit.

In 2010 maten wij de vooruitgang in video intelligentie door een videozoekmachine uit 2006 te vergelijken met een uit 2009 (Snoek & Smeulders, 2010). In slechts drie jaar tijd, waren de prestaties van de zoekmachine verdubbeld. Het begon te werken. Zelfs zo goed dat de eerste startups op de markt begonnen te verschijnen die de software aan de man probeerde te brengen, waaronder het Amsterdamse Euvision. Daarmee kunnen we concluderen dat video intelligentie werkt. We kunnen boom, roos, vis en alle andere zelfstandige naamwoorden lezen uit video.

2 De kinderjaren voorbij

En toen kwam er een doorbraak, die ik en vele met mij niet eerder hadden meegemaakt. Het jaar 2012 was de grote doorbraak van diepe neurale netwerken (Goodfellow *et al.*, 2017), een techniek waar al aan gesleuteld werd sinds het begin van de Tweede Wereldoorlog. Losjes geïnspireerd op het menselijke

brein tracht een neuraal netwerk de verbindingen tussen neuronen en synapsen te leren die de beste voorspelling voor een sensorische input opleveren. Waar de neurale netwerken in het verleden faalden om de hooggespannen verwachtingen waar te maken, werden de verwachtingen dit keer overtroffen. Geholpen door grote verzamelingen voorbeelden om van te leren, snelle grafische GPU-kaarten voor het rekenwerk en het vermogen om lagen van netwerken te stapelen, won deep learning alle competities in de spraakherkennig, de beeldherkenning, de gezichtsherkenning en vele andere vakgebieden. Deep learning versloeg zelfs de wereldkampioen in het bordspel Go. Ook voor de video intelligentie was de impact ongekend. We maakten als vakgebied weer een enorme sprong vooruit.

Maar niet alleen voor het lezen van zelfstandig naamwoorden, ook voor bijvoegelijke naamwoorden, werkwoorden en zelfs hele zinnen is aangetoond dat ze tot op zekere hoogte volautomatisch herkend kunnen worden in video. We weten niet alleen of iets in een video gebeurt maar ook waar het gebeurt. Daarmee wordt ook tellen in videobeelden een gemakkelijke klus. Door deze en andere verworvenheden staat het veld in de warme belangstelling. De belangrijkste conferentie van het veld groeide van krap 2000 bezoekers in 2012 tot meer dan 6000 bezoekers bij de laatste editie in juni van dit jaar. Er werden vorig jaar 3500 artikelen ingestuurd, dit jaar 5000. En de rek is er nog niet uit. Dit komt vooral door de industrie die zich *en masse* op deep learning en beeldherkenning hebben gestort. Aangetrokken door de belofte en de overspannen verwachtingen, worden startups ingelijfd, talent weggekocht en is video intelligentie en kunstmatige intelligentie in het algemeen een van de heetste topics in Silicon Valley, in China en sinds kort ook in Europa.

Het moge duidelijk zijn, video intelligentie staat op het punt de kinderjaren achter zich te laten en dan volgt onvermijdelijk de puberteit. Daar hoort ook het verkennen van maatschappelijke en ethische grenzen bij. Video intelligentie wordt bijvoorbeeld ingezet voor het genereren van video's die moeilijk van echt zijn te onderscheiden, met als favoriete applicaties wraakporno en roddelen. Met software die u gratis van het internet plukt kunt u nu iedereen van alles laten doen en zeggen, door alleen maar een paar voorbeeld filmpjes van diens gezicht te verzamelen. U heeft de jolige filmpjes met Donald Trump, Vladimir Poetin en Mark Rutte vast wel eens voorbij zien komen[4]. Wat we zien en horen in een video hoeft geen waarheid te zijn. Onderzoekers zijn momenteel in een wedloop verwikkeld, waarbij de ene helft zo goed mogelijk valse video's probeert te genereren en de andere helft het als haar taak ziet die nepvideo's te ontmaskeren. Als een gevolg hiervan zullen deze nepvideo's steeds realistischer worden. Het is maar dat u het weet.

Video intelligentie zal in de nabije toekomst ook het gedrag van zelfstanding opererende machines bepalen. Het bekendste voorbeeld is de zelfrijdende auto. Deze zal leiden tot minder dodelijke slachtoffers in het verkeer, daar is elke expert het over eens. Maar de slachtoffers die de zelfrijdende auto wel maakt, zouden waarschijnlijk niet gemaakt worden door menselijke chauffeurs. Recente dodelijke ongelukken van Tesla's en Uber's zijn hier treurige voorbeelden van[5]. Hoe hiermee om te gaan? Een ander heet hangijzer is de inzet van video intelligentie voor defensieve doeleinden. Personeel van Google was onlangs woedend toen bekend werd dat de zoekmachine grootmacht meewerkte met het Amerikaanse ministerie van defensie om in dronebeelden Afghaanse voertuigen, mensen en gebouwen te herkennen[6]. Pubers hebben huisregels nodig, het moment is daar om die ook voor video intelligentie, en de machines die er gebruik van maken, af te spreken. Dat is een gezamenlijke verantwoordelijkheid van wetgever, technologen en maatschappij.

In '1984' schetste George Orwell een dystopische wereld die gekenmerkt wordt door fake nieuws, grootschalige surveillance en eeuwigdurende oorlog (Orwell, 1948). We zijn gelukkig nog niet zo ver, maar we zijn wel enigszins op weg. Eigenaarschap en gebruik van publieke videodata moet beter geregeld worden. In de VS is de tech-industrie de eigenaar van deze data, in China is het de overheid, in Europa zien we het liefste het individu als de rechthebbende[7]. Het zal niet meevallen om dit goed te regelen, maar wellicht kunnen we inspiratie halen uit de geschiedenis. Zo'n 150 jaar geleden werd de fotografie volwassen en dat leidde ook tot misbruik, ophef en vertier. Het duurde even voor de maatschappij er een passend antwoord op had, maar in 1890 publiceerde Warren en Brandeis hun beroemde artikel 'The Right to Privacy' (Warren & Brandeis, 1890) waarin ze uiteenzetten dat elk mens het recht heeft om met rust te worden gelaten. Het privacyrecht was geboren. Van recentere datum is het recht om vergeten te worden, dat ontstond toen het Europese Hof van Justitie in 2014 besloot dat iedere Europeaan het recht moet hebben bepaalde resultaten uit een zoekmachine te laten verwijderen. Het recht om onzichtbaar te zijn in publiek toegankelijke video lijkt een logische volgende stap. Technologisch moet dat in ieder geval haalbaar kunnen zijn.

3 Op weg naar onafhankelijkheid

Waar staan we vandaag? Video intelligentie die leert van voorbeelden werkt als een tierelier, maar alleen voor taken waarvoor slechts een beperkt begrip van de wereld noodzakelijk is. Voor complexere vraagstukken moeten we nog

veel leren. Men neme het voorbeeld van de fietsendief (zie figuur 4). Een echt Amsterdams probleem. Dit vereist een tamelijk specifieke analyse van het beeld, waarbij de nuance zit in het slot losmaken of openbreken. Ik verwacht dat dit met voldoende voorbeelden misschien kan lukken, maar die voorbeelden zijn niet ruim voorhanden, de dader laat zich immers niet graag zien.

Figuur 4 **Een Amsterdams probleem. Wordt het slot van de fiets geopend of gebroken? De nuance is lastig om automatisch te kunnen begrijpen, vooral als voorbeelden om van te leren ontbreken.**

Dit plaatst ons voor een paradox. Namelijk, hoe preciezer het begrip van videointelligentie dat noodzakelijk is, hoe onrealistischer de aanname dat voldoende voorbeelden gevonden kunnen worden om van te leren. En nog erger, hoe preciezer het moet zijn, hoe meer werk juist het verzamelen van die voorbeelden wordt. We willen niet alleen weten wat er globaal in een video gebeurt, maar van elk beeldpunt achterhalen bij welk object het hoort, waar die objecten mee interacteren en wat dat gedurende de video betekent. De missie van het lab voor de komende jaren is daarom duidelijk en simpel, op weg naar een compleet begrip van video met zo min mogelijk voorbeelden.

Hoe daar te geraken? De inspiratie komt vanuit Nobelprijswinnaar Philip Warren Anderson. Hij schreef in 1972 het artikel 'More is Different' (Anderson, 1972) waarin hij uiteenzette dat het wetenschappelijke streven naar reductionisme beperkingen heeft, er zijn hiërarchieën in de wetenschap voor een reden, elk met zijn eigen fundamentele principes voor vooruitgang. Niet elk fenomeen kan beschreven worden door een natuurwet, niet alles kan geleerd worden van voorbeelden, of vrij naar Andersson: *Video is different.*

Video heeft een aantal cruciale eigenschappen die verrassend genoeg maar beperkt benut worden binnen de state-of-the-art in video intelligentie. Op de eerste plaats: *tijd*. Voor veel bestaande methoden maakt het niet uit of je de video vooruit of achteruit afspeelt, ze zullen altijd hetzelfde antwoord opleveren. Dat is op zijn minst vreemd. Tijd heeft nog een voordeel. Je kunt het gebruiken als een signaal om van te leren. Dat is precies wat we nu aan het doen zijn. Op de tweede plaats: *ruimte.* Bestaande methoden reduceren een video tot een set van losse afbeeldingen en detecteren de locatie van objecten

en activiteiten in iedere afbeelding afzonderlijk en onafhankelijk. Dat is zonde omdat de natuurlijke redundantie van video genegeerd wordt. Op de derde plaats: *multimodaliteit.* Een video is meer dan beeld. Geluid, spraak en tekst bevatten waardevolle informatie die nog maar beperkt gebruikt worden. Dit terwijl ze waardevol zijn voor human-computer interactie met video en wederom een waardevolle bron zijn om van te leren, vaak gratis en voor niets. Om de missie van het lab te realiseren is het dus van belang om video ook daadwerkelijk als video te behandelen.

Dat hoef ik gelukkig niet alleen te doen en ik prijs mezelf gelukkig te mogen samenwerken met een team van internationale talenten in diverse onderzoeksprojecten. Ik wil een drietal projecten kort noemen.

Het grote Amerikaanse Qualcomm nam met de aankoop van spin-off Euvision in 2014 een risico door zich in Amsterdam te vestigen en een samenwerking aan te gaan met de Universiteit. Die publiek-private samenwerking werd het QUVA Lab, waar we naast video intelligentie werken aan beeldherkenning en machine leren voor mobiele toepassingen. De samenwerking met Qualcomm vormde ook de opmaat voor meer en heeft in Amsterdam geleid tot het Innovatie Centrum voor Kunstmatige Intelligentie. Een nationaal initiatief dat zich als doel heeft gesteld om kunstmatige intelligentie te ontwikkelen in samenwerking met de industrie, door middel van gezamenlijke onderzoekslabs. Collega-hoogleraren openden al labs met Bosch, Ahold-Delhaize, Elsevier en de nationale Politie, er zullen er nog vele volgen.

Op Schiphol hangen meer dan 3500 camera's die het terrein binnen en buiten 24 uur per dag monitoren. Het bekijken van deze videobeelden gebeurt nu nog door beveiligingspersoneel, douanebeambten en de Koninklijke Marechaussee. Samen met Schiphol, TNO en de Vrije Universiteit van Amsterdam, gaan we dit proces automatiseren met video intelligentie. In het bijzonder door het automatisch herkennen van gepast en ongepast gedrag, zoals mensen die flauwvallen, en het volgen van zakkenrollers over meerdere camera's. De financiering komt vanuit het nationale NWO perspectief programma rond 'Efficient Deep Learning'.

De opvolger van Euvision heet Kepler Vision, binnen deze startup werken we aan commercialisatie van video intelligentie. We specialiseren in het herkennen van lichaamstaal in al haar facetten, met mogelijke applicaties in de zorg, de media en het verkeer. Hoewel we nog maar net zijn begonnen is het team er al in geslaagd om begin januari haar technologie te mogen demonstreren in het Holland Paviljoen op de grote Consumer Electronic Show in Las Vegas, aan mij de eer om u hier vandaag alvast een eerste voorproefje te geven. De software monitort of iemand wel voldoende eet en drinkt of in een sociaal isolement aan het raken is. Ook kan de software onderscheid maken

tussen iemand die op de grond ligt – en dus gevallen is – en iemand die op de bank ligt om uit te rusten. De software is bedoeld om zorgmedewerkers te ondersteunen bij hun surveillerende taak (zie figuur 5).

Figuur 5 **Demonstratie van Kepler Vision's technologie voor het herkennen van lichaamstaal, waarbij onder meer te zien is hoeveel tijd geobserveerde personen met bepaalde activiteiten bezig zijn geweest.**

4 Een leven lang leren

Video intelligentie zal ook impact hebben in het onderwijs. In China lopen ze voorop. Een middelbare school in Hangzhou heeft onlangs een klaslokaal uitgerust met camera's die registreren of scholieren lezen, schrijven, handen opsteken, etc[8]. Ook lezen ze de emoties van de leerlingen. De camera's zien of iemand blij, verdrietig, boos of juist verveeld is. Op deze manier krijgt de docent direct feedback of zijn of haar lessen aankomen. Ook wordt de docent ondersteund bij het monitoren van het welzijn van individuele leerlingen. Een ander voorbeeld is surveilleren bij tentamens. Als hoogleraar word ik ook geacht hier mijn bijdrage aan te leveren. Uiteraard wordt dit naar goed academisch gebruik uitbesteed aan postdocs en promovendi, maar dit kan net zo

goed, of zelfs beter, met video intelligentie. Om deze en alle andere dromen te realiseren zijn slimme meiden en jongens nodig die kunstmatige intelligentie willen leren, de materie uiteindelijk begrijpen en die er mee kunnen innoveren. Onderwijs in kunstmatige intelligentie is daarbij de cruciale stap voorwaarts.

Met onderwijs kun je niet vroeg genoeg beginnen. Ook dat hebben ze in China uitstekend begrepen. Confucius zei het al: 'Als je een plan voor een jaar hebt, plant dan rijst; als je een plan voor een decennium hebt plant dan bomen, en als je een plan voor het leven hebt onderwijs dan je kinderen.' Het Chinese boek 'Fundamenten van Kunstmatige Intelligentie' voor middelbare scholieren werd dit jaar geïntroduceerd (Tang & Chen, 2018). Volgend jaar al volgt een hele serie van maar liefst tien volumes voor zowel basis- als voortgezet onderwijs. Op honderden Chinese scholen zal het dit voorjaar als keuzevak worden aangeboden of onderdeel uitmaken van de vaste stof. Ook in het Verenigd Koninkrijk heeft de overheid, ondanks de Brexit perikelen, een plan aangekondigd om samen met mini-computermaker Raspberry Pi veertigduizend onderwijzers uit het basis- en voortgezet onderwijs te trainen in het geven van informatica-onderwijs[9]. Nederland faalt tot op heden om overtuigend kunstmatige intelligentie en informatica te onderwijzen in het basis- en voortgezet onderwijs. De Koninklijke Nederlandse Academie van Wetenschappen observeerde dit reeds in 2012 en concludeerde 'De huidige vakken Informatiekunde en Informatica op havo en vwo hebben een marginale positie, schieten kwalitatief tekort en zijn inhoudelijk uit de tijd.'[10] Het moment is aangebroken om daar iets aan te doen, ik wil me daar graag voor inzetten en doe bij deze een oproep aan mijn collega-hoogleraren om me daarbij te helpen. Om er ook een Nederlands gezegde tegen aan te gooien, wie de jeugd heeft, heeft de toekomst.

Ondanks de beperkte aandacht voor kunstmatige intelligentie in hun vooropleiding, kiezen gelukkig steeds meer middelbare scholieren juist voor deze studie. De gevolgen daarvan heeft u van de zomer misschien voorbij zien komen in het nationale nieuws. Opleidingen worden overspoeld met aanvragen en kunnen de druk simpelweg niet aan. Zo ook bij onze eigen opleiding aan de Universiteit van Amsterdam, waar we het aantal aanmeldingen voor de master zagen verdubbelen van 300 tot 700[11]. Hoe om te gaan met deze populariteit? Het antwoord is vooralsnog de rem erop door numerus fixus en selectiecriteria. Dat is onwenselijk, de maatschappij en het bedrijfsleven heeft de expertise juist nu hard nodig. Opschalen door het inhuren van extra personeel is lastig gebleken omdat de mensen die over zowel de onderzoeks- als onderwijsexpertise beschikken niet ruim voorhanden zijn. Bovendien kunnen zij voor een veel aantrekkelijker salaris ook terecht in het bedrijfsleven.

De roep om extra financiële middelen vanuit de overheid is groot, maar wat doen we ermee als we het straks krijgen? Sommigen geloven dat we talent van buiten aan kunnen trekken, maar dat lijkt mij een ijdele hoop. In Engeland, Duitsland en Frankrijk hebben ze hetzelfde plan. We kunnen beter investeren in het talent dat al binnen is, de hoogleraren van morgen, die nu al een groot deel van het onderwijs voor hun rekening nemen. Hun tijd is kostbaar. Laat de universitair docenten tijd winnen door ze voor onderwijs te belonen met promovendi, zodat ze niet in de knel komen met hun onderzoek. Ondersteun ze in de administratieve processen van de universiteit, een secretaresse is gemakkelijker in te huren dan een machine learning hoogleraar. En hoewel salaris geen rol zou moeten spelen voor de keuze van een academische carrière is het moment wel daar om eens na te denken over een gepaste marktconforme toelage voor postdocs en universitair docenten in de kunstmatige intelligentie. Al was het alleen maar om de symboliek die ervan uitgaat.

Voor de mensen in de zaal die nu ook meer willen weten over kunstmatige intelligentie, maar geen zin hebben in een nieuwe studie, heb ik goed nieuws. Op 21 december wordt er een online AI-cursus voor kunstmatig intelligentie gelanceerd[12], in het Nederlands, waarbij u gratis en in uw eigen tempo de stof tot u kan nemen. Ik raad het u van harte aan.

5 Wat heeft u geleerd?

Met de aankondiging van het huiswerk, zijn we bijna aan het einde beland van deze rede. Wat heeft u geleerd van dit college? Ik heb u duidelijk willen maken dat elk voorwerp met plaats voor een camera en de potentie voor een toepassing er een zal krijgen. Camera's moeten daarom niet alleen video creeren maar deze beelden ook volautomatisch interpreteren. De technologie die dat mogelijk maakt heet video intelligentie en deze staat op het punt de kinderjaren achter zich te laten. Net als elke andere technologie heeft video intelligentie geen geweten. Voor het bewaken van maatschappelijke en ethische grenzen zijn en blijven we met zijn allen zelf verantwoordelijk. De wetenschap heeft daarbij een belangrijke signalerende taak en kan helpen bij het aandragen van mogelijke oplossingen en inzichten. Nog belangrijker is het opleiden van de nieuwe generatie die kan lezen en schrijven met kunstmatige intelligentie, en daar kunnen we niet vroeg genoeg mee beginnen. Afhankelijkheid van voorbeelden is nog een achilleshiel voor video intelligentie, compleet begrip vereist nog jaren onderzoek. Een troostende gedachte.

Woord van dank

Graag wil ik nog een aantal mensen bedanken. Op de eerste plaats het College van Bestuur van de Universiteit van Amsterdam, evenals het bestuur van de Faculteit der Natuurwetenschappen, Wiskunde en Informatica, in het bijzonder de decaan, voor het in mij gestelde vertrouwen.

Promotor Arnold Smeulders, die mij na het voltooien van mijn proefschrift op een dag een project noemde. Ik kon destijds niet precies bevroeden wat dat betekende maar ik wist wel dat het met projecten van Arnold altijd goed komt. Arnold, bedankt dat je me al die jaren een beeld hebt voorgehouden: soms een visioen, meestal een voorbeeld, vaker nog een spiegel. Ik hoop dat we ook na je emeritaat nog een aantal jaren kamer- en reisgenoten blijven.

Arnold is van het beeld, co-promotor Marcel Worring is meer van de wortel. Die moet je mensen ook voorhouden dan gaan ze heel hard rennen. Marcel bedankt dat je mijn wetenschapsmotortje aanzwengelde, zorgde voor de eerste successen met de MediaMill videozoekmachine en er uiteindelijk als instituutsdirecteur ook voor hebt gezorgd dat ik die wortel vandaag kan grijpen.

Hoe fantastisch is het dat ik de onderzoeksgroep waar ik wetenschappelijk ben opgegroeid nu zelf mag leiden? Ik prijs me daarbij gelukkig met de steun van Dennis Koelma, Virginie Mes en sinds dit jaar Petra Venema die o zo belangrijk zijn voor de sfeer en temperatuur in het lab, soms ook letterlijk als de GPU's oververhit raken door het vele rekenwerk. Stratis Gavves wil ik danken voor zijn aanstekelijke enthousiasme, nieuwsgierigheid en zijn leidende rol bij het QUVA lab. Het is mooi om te zien hoe je groeit en bloeit. Dat geldt zeker ook voor alle promovendi en postdocs, jullie zijn het lab, laten we er samen iets heel moois van maken.

Huidige en voormalige collega's van Euvision, Qualcomm en Kepler wil ik danken voor de fijne samenwerking en hun vermogen om wetenschap om te zetten in een echte toepassing. Daar is een grandioze inspanning voor nodig die door velen in de wetenschap wordt onderschat. Harro Stokman snapt dat als geen ander en is ook nog eens in staat om er een winstgevend bedrijf omheen te bouwen. Harro, onder jouw leiding wordt Kepler ongetwijfeld ook een daverend succes. No pressure.

Gelukkig is er ook nog een leven naast video intelligentie. Het is fijn om altijd terug te kunnen vallen op vrienden, familie en schoonfamilie. Speciaal woord van dank voor mijn ouders Paul en Woltje; voor jullie steun, aanmoediging en levenshouding. Ik hoop dat ik die kan doorgeven aan mijn eigen kinderen. Die zijn nu nog klein en zo'n jong gezin is vaak best wel hectisch, maar gelukkig kan altijd een beroep worden gedaan op oma Aal om rust in de

tent te houden. Tot slot wil ik Marga danken voor haar steun, begrip en engelengeduld. Lest best.

Ik heb gezegd.

Noten

1. LDV Capital. '45 Billion Cameras by 2022 Fuel Business Opportunities.' 12 augustus, 2017.
2. Klaassen. N., 'Politie krijgt steeds vaker beelden privécamera.' Algemeen Dagblad, 27 juni, 2018.
3. NIST TRECVID. https://trecvid.nist.gov
4. Romano, A., 'Jordan Peele's simulated Obama PSA is a double-edged warning against fake news.' Vox, 18 april, 2018.
5. Death of Elaine Herzberg. https://en.wikipedia.org/wiki/Death_of_Elaine_Herzberg
6. Shane, S., Metz, C. & Wakabayashi, D., 'How a Pentagon Contract Became an Identity Crisis for Google.' The New York Times, 30 mei, 2018.
7. Voor discussie zie: Keen, A., *How to Fix the Future.* Atlantic Books, London, 2018.
8. People's Daily. 'Facial recognition used to analyze students' classroom behaviors.' 19 mei, 2018.
9. Murgia, M., 'How the UK plans to teach computer science to every child.' Financial Times, 27 november, 2018
10. Koninklijke Nederlandse Akademie van Wetenschappen. 'Digitale geletterdheid in het voortgezet onderwijs.' December 2012.
11. Het Parool. 'UvA worstelt met grote groei studie Kunstmatige Intelligentie.' 16 juli, 2018.
12. De Nationale AI cursus. https://www.ai-cursus.nl/

Referenties

Anderson, P.W., 'More Is Different.' In: *Science*, vol. 177, no. 4047, p. 393-396, 1972

Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*. MIT Press, Cambridge, MA, 2017

Maxmen, A., 'Hidden lives of deep-sea animals.' In: *Nature,* vol. 561, p. 296-297, 2018

Orwell, G. *1984*. Samaira Book Publishers, 2017

Parker, J., *In the Blink of an Eye.* The Trustees of the National History Museum, London, 2016

Snoek, C.G.M. & Smeulders, A.W.M., 'Visual-Concept Search Solved?' In: *IEEE Computer*, vol. 43, no. 6, p. 76-78, 2010

Tang, X. & Chen, Y., 人工智能基础 *[Fundamentals of Artificial Intelligence]*. East China Normal University Press, 2018

Wactlar, H. D., Kanade, T. Smith, M. A., & Stevens, S. M., 'Intelligent access to digital video: Informedia project.' In: *IEEE Computer*, vol. 29, no. 5, p. 46-52, 1996

Warren, S.D, & Brandeis, L.D., 'The Right to Privacy.' In: *Harvard Law Review*, vol. 4, no. 5, p. 193-220, 1890