

Kimi af Forselles

Taustat konenäön hyödyntämisestä markkinointiin

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Helmikuu 2020

TIIVISTELMÄ

Kimi af Forselles: Taustat konenäön hyödyntämisestä markkinointiin
Kandidaattitutkielma
Tampereen yliopisto
Tietojenkäsittelytieteiden tutkinto-ohjelma
Helmikuu 2020

Tämä tutkielma esittelee konenäön keinoja tulkita visuaalista dataa sekä sen mahdollisuuksia markkinoinnissa. Tutkielma alkaa konenäön määrittelyllä ja rinnastuksella ihmisen näköjärjestelmään. Konenäön määrittelyn jälkeen avataan neuroverkkojen perusarkkitehtuuria ja oppimismekanismeja yksinkertaisen konenäön sovelluksen avulla. Neuroverkkojen jälkeen käsitellään konenäön tasoa tällä hetkellä erinäisten mittareiden avulla sekä pohditaan konenäön haasteita kuvien syvällisemmässä ymmärryksessä. Lopuksi käydään vielä läpi keinoja, miten konenäköä voidaan hyödyntää markkinoinnissa.

Avainsanat: tekoäly, konenäkö, markkinointi, neuroverkot

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Sisällys

1	Johdanto.....	1
2	Konenäön määrittely.....	2
2.1	Konenäkö ja ihmisen visuaalinen järjestelmä	2
2.2	Neuroverkot.....	3
3	Konvoluutioverkot ja suunta kohti kuvien syvempää ymmärrystä	5
4	Yhteenveto.....	11
5	Viiteluettelo	13

1 Johdanto

Digitalisoitumisen myötä informaation määrä on kasvanut eksponentiaalisesti. Domon [2018] teettämän arvion mukaan vuonna 2020 uutta informaatiota syntyy 1,7 megatavua jokaista maapallon ihmistä kohti joka sekunti. Päivässä uutta dataa syntyy siis $6,7 \cdot 10^{20}$ tavua, eli noin saman verran kuin maapallolla on hiekanjyviä [Krulwich 2012]. Cisco Systemsin teettämän tutkimuksen mukaan vuonna 2022 jopa 82 % internetin liikenteestä on videodataa [Cisco VNI 2018]. Visuaalista dataa on saatavilla valtava määrä, mutta tietokoneille sen tulkinta on vaikeaa. Ihmisen aivot ovat sen sijaan äärimmäisen hyviä prosessoimaan visuaalista dataa ja tulkitsemaan kuvia [Potter *et al.* 2014]. Ihminen voi jo nopean silmäyksen perusteella kuvailla näkemäänsä kuvaa erittäin tarkasti. Ihmisille nähdyn ymmärtäminen on itsestäänselvyys, eikä konenäönkään aluksi uskottu olevan vaikeaa. Konenäköä alettiin tutkia tekoälyn kultavuosina 60-luvulla ja vilsimmät ennustukset olettivat tietokoneen älykkyyden olevan ihmisen tasolla jo muutaman vuosikymmenen kuluttua. Pian kuitenkin huomattiin, ettei tekoälyä eikä konenäköä ratkaistakaan niin helposti. Konenäkö onkin kamppailut kuvien ymmärtämisen kanssa vuosikymmeniä ja vasta viime vuosina kehittynyt sellaiselle tasolle, että sitä voidaan hyödyntää reaali maailman ongelmissa [Chen 2016].

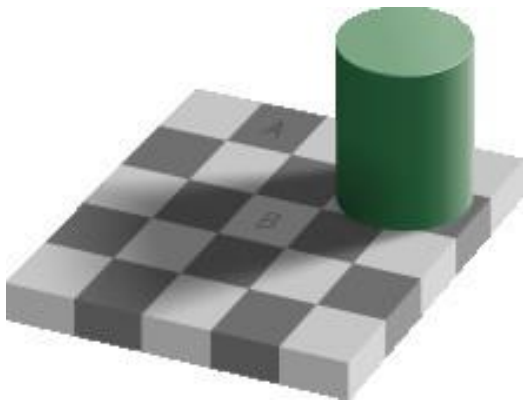
Markkinoinnissa konenäköä voidaan hyödyntää eritoten kontekstuaalisessa mainonnassa. Kontekstuaalisella mainonnalla tarkoitetaan mainosten näyttämistä relevanteissa paikoissa, esimerkiksi lemmikkikaupan mainos ennen koirankoulutusvideota YouTubessa. Internetin myötä mainontaa on kohdennettu keräämällä henkilökohtaista dataa evästeiden kautta, mutta viime aikoina yksityisyys on noussut pinnalle: yleinen tietosuoja-asetus GDPR rajoittaa henkilökohtaisen datan käyttöä. Ihmiset käyttävät myös enemmän ja enemmän internetselainten inkognito-tiloja, jotka eivät tallenna evästeitä. Myös yli 60 % internetin selailusta tapahtuu mobiilissa, joka entisestään vaikeuttaa henkilökohtaisten evästeiden keruuta [GumGum 2019]. Mobiilissa käytetään paljon sovelluksia, joiden evästeet eivät siirry apista toiseen vaikeuttaen käyttäjän tottumusten tulkintaa. Konenäkö tarjoaa tähän ratkaisun analysoimalla itse verkkosivujen sisältöä ja asettamalla mainokset relevanttiin, bränditurvalliseen kontekstiin.

Seuraavaksi luvussa 2 käydään läpi hieman konenäön historiaa, sekä avataan neuroverkkojen toimintaperiaatetta. Luvussa 3 esitellään konenäön kehitystä lähivuosina, sekä pohditaan konenäön haasteita tulevaisuudessa. Lopuksi neljännessä luvussa tiivistetään, miksi konenäkö on markkinoinnille tärkeä työkalu, sekä miten se voidaan valjastaa markkinoinnin käyttöön.

2 Konenäön määrittely

2.1 Konenäkö ja ihmisen visuaalinen järjestelmä

Konenäkö on visuaalisen datan prosessointia ja tulkintaa. Konenäön avulla pyritään luomaan laskennallisia malleja ihmisen visuaalisesta järjestelmästä ja muodostamaan itsenäisiä systeemejä suorittamaan samoja tehtäviä mitä ihminenkin voi tehdä [Huang 1996]. Ihmisen näköjärjestelmässä verkkokalvo muuntaa ensin kolmiulotteisen maailman kaksiulotteiseksi ja aivot muuntavat maailman takaisin kolmiulotteiseksi. Silmän neuronit eivät kuitenkaan lähetä aivoille täsmälleen kaikkea verkkokalvolle heijastunutta tietoa, vaan painottavat suuria valon intensiteetin muutoksia ja reunoja. Havaituista piirteistä muodostetaan sisäisiä malleja vastaava järkeenkäypä kuva hyödyntäen esimerkiksi stereonäköä, varjoja, tekstuuria ja esineiden suhteellista kokoa [Marr 2010]. Nämä sisäiset mallit voivat johtaa erinäisiin havainnointiharhoihin, esimerkiksi tummassa ympäristössä värit näyttävät kirkkaammilta ja kirkkaassa ympäristössä tummemmilta (Kuva 1).

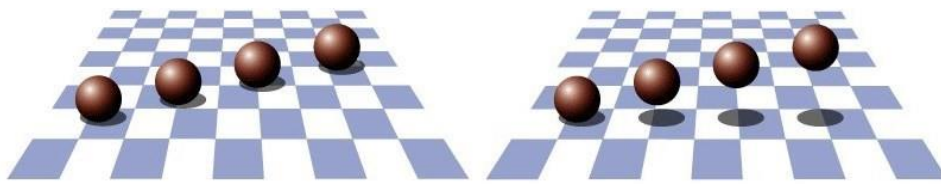


Kuva 1. “Checker-shadow illusion”. Kohdat A ja B ovat samanvärisiä [Anderlson 1995].

Stereonäkö muodostuu, koska ihmisellä on kaksi silmää. Silmät ovat hiukan erillään toisista, joten molempiin silmiin tulee hieman eri kuva. Tämän avulla aivot pystyvät muodostamaan 2D-kuvasta 3D-kuvan [Marr 2010]. Perinteiset 3D-lasit käyttävät tätä samaa tempua: toinen linssi suodattaa punaisen värin pois, toinen vihreän, jolloin molemmat silmät näkevät eri kuvan ja 2D-kuva ilmenee kolmiulotteisena.

Aivot hyödyntävät myös varjoja muodostaessaan kolmiulotteista kuvaa. Valon oletetaan tulevan ylhäältä, jolloin varjon sijainnin avulla voidaan vaikuttaa kappaleen havaittuun korkeuteen [Serendip 1994]. Kuvassa 2 on identtiset laudat ja neljä palloa samassa

paikassa. Varjon sijainnin muuttuessa aivot tulkitsevat pallojen korkeuden muuttuneen [Grobstein 2003].



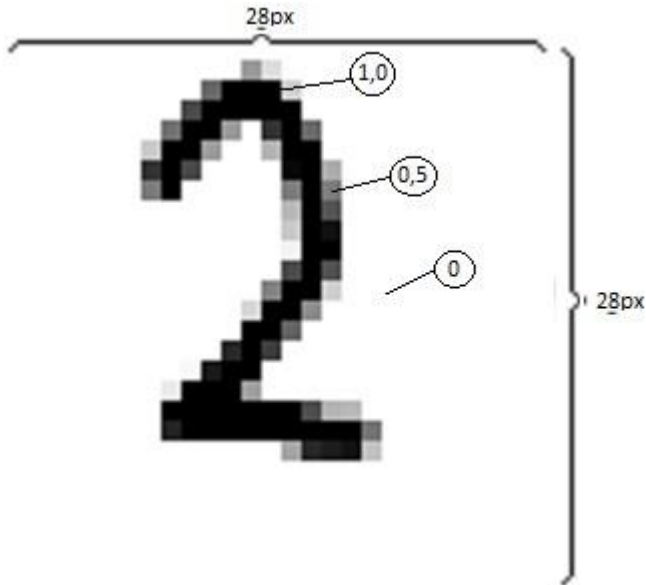
Kuva 2. ”Ball and shadow illusion” [Serendip 1994].

Kolmiulotteisesta kappaleesta voi muodostaa loputtoman määrän erilaisia kaksiulotteisia esityksiä kappaleen valotusta, asentoa ja etäisyyttä muuntamalla [Pinto *et al.* 2008]. Roberts [1963] lähti tutkimaan konenäköä juuri tämän ongelman kautta: ennen kuin kappaleet voidaan tunnistaa, tulee niistä luoda 3D-esitys. Hän pyrki tunnistamaan yksinkertaisten kappaleiden reunat ja luomaan niistä kolmiulotteisen mallin, jota voi tarkastella eri kulmista [Roberts 1963]. Kappaleiden tunnistuksen ei uskottu olevan vielä vaikeaa, joten legendan mukaan Marvin Minsky antoi oppilailleen kesätyöksi tehtävän opettaa tietokone kuvailemaan näkemäänsä [Papert 1966]. Ensin pyrittiin eristämään kuvasta kappaleet ja tausta ja sitten tunnistamaan ja nimeämään kappaleet. Konenäkö osoittautui kuitenkin erittäin haastavaksi ongelmaksi, jota on tutkittu jo vuosikymmeniä.

2.2 Neuroverkot

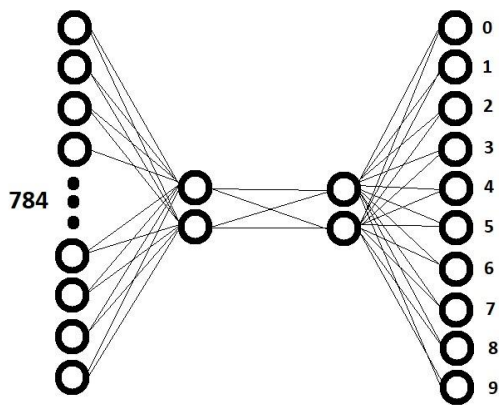
Neuroverkko koostuu neuroneista, jotka ovat kerroksittain yhteydessä toisiinsa. Jokaisen yhteyden välillä on painoarvo ja jokaisella neuronilla on kynnysarvo. Näitä arvoja säätämällä verkko oppii. Neuroverkon toimintaperiaate havainnollistuu parhaiten esimerkein: seuraavaksi käydään läpi neuroverkon rakenne käsinkirjoitettuja numeroita tunnistettaessa [Nielsen 2015].

Aluksi kuva esikäsitellään 28x28 pikselin kokoiseksi. Tämän jälkeen jokainen pikseli saa arvon nollan ja yhden välillä, riippuen pikselin väristä (Kuva 3).



Kuva 3. Käsini kirjoitettu numero 2 esikäsiteltynä.

Neuroverkon ensimmäisessä kerroksessa on 28x28, eli 784 neuronit. Nämä neuronit saavat syötteenä kuvan jokaisen pikselin arvon. Neuroverkon viimeisessä kerroksessa on 10 neuronit, joiden aktivaatio kuvaa minkä numeron neuroverkko sai syötteenä. Ensimmäisen ja viimeisen kerroksen välissä on x määrä piilokerroksia (hidden layer).



Kuva 4. Esimerkki neuroverkko. Todellisuudessa piilokerroksissa on paljon enemmän neuroneja.

Kuvassa 4 on nähtävissä yksinkertainen neuroverkon rakenne: jokainen neuronit on yhteydessä seuraavan kerroksen jokaiseen neuroniin ja jokaisen yhteyden välillä on painoarvo. Neuronit saavat syötteenä kaikkien edellisen tason neuronien painotetun summan. Summa syötetään aktivaatiofunktioon, esimerkiksi jos summan halutaan olevan yhden ja nollan välillä, asetetaan se sigmoid-funktioon. Neuronilla on myös kynnsarvo, joka määrittää, koska neuronit aktivoituu. Kynnsarvo lisätään painotettuun summaan ennen aktivaatiofunktioita. Jokainen neuronit syöttää uuden arvonsa eteenpäin, kunnes

viimeisessä kerroksessa neuronin aktivaatio kuvaa, minkä käsinkirjoitetun numeron verkko sai syötteenä.

Painot ja kynnsarvot ovat kuin nuppeja, joita säätämällä verkko oppii. Parametreja on kuitenkin valtava määrä, joten niiden manuaalinen säätö on mahdotonta. Neuroverkko koulutetaan syöttämällä sille dataa ja vertailemalla saatuja tuloksia oikeisiin tuloksiin. Käytännössä tämä voidaan toteuttaa laskemalla yhteen haluttujen arvojen ja saatujen arvojen neliöiden erotukset. Tämä antaa jokaisen koulutusesimerkin kohdalla arvon, joka kuvaa miten hyvin neuroverkko suoriutui. Kaikkien koulutusesimerkkien kohdalla tehdään sama ja siitä otetaan keskiarvo. Jotta neuroverkko voi saavuttaa täyden potentiaalinsa, tulee löytää nappien asento, joilla tämä keskiarvo on pienin mahdollinen. Tämä toteutetaan laskemalla jokaiselle parametrille gradienttivektori, joka kertoo mihin suuntaan parametria tulee muuttaa, jotta saadaan pienempi arvo. Tätä prosessia toistetaan, kunnes löydetään paras arvo.

Neuroverkot voivat koulutuksen yhteydessä ylioppia tai alioppia. Ylioppimisella tarkoitetaan sitä, kun neuroverkko on hyvä tunnistamaan harjoitussetinsä dataa, muttei osaa tunnistaa uutta dataa. Alioppimisella tarkoitetaan sitä, kun neuroverkko ei opi tunnistamaan edes harjoitusdataa. [Chen 2016] Alioppiminen voidaan eliminoida esimerkiksi lisäämällä neuronien kerroksia ja neuronien määrää sekä parantamalla harjoitusdatan määrää ja laatua. Ylioppimista voidaan välttää kouluttamalla verkko uudestaan eri lähtöpainoarvoilla tai kouluttaa samaa neuroverkkoa useilla eri lähtöpainoarvoilla rinnakkain. Mitä enemmän yhteyksiä neuroverkossa on, sitä enemmän koulutusesimerkkejä vaaditaan, jotta voidaan välttää ylioppimista. Ratkaisuksi tähän on kehitetty suuria, valmiiksi luokiteltuja datajoukkoja, kuten ImageNet, jossa on yli 14 miljoonaa kuvaa ja 20 000 kategoriaa [Deng et al. 2009].

3 Konvoluutioverkot ja suunta kohti kuvien syvempää ymmärrystä

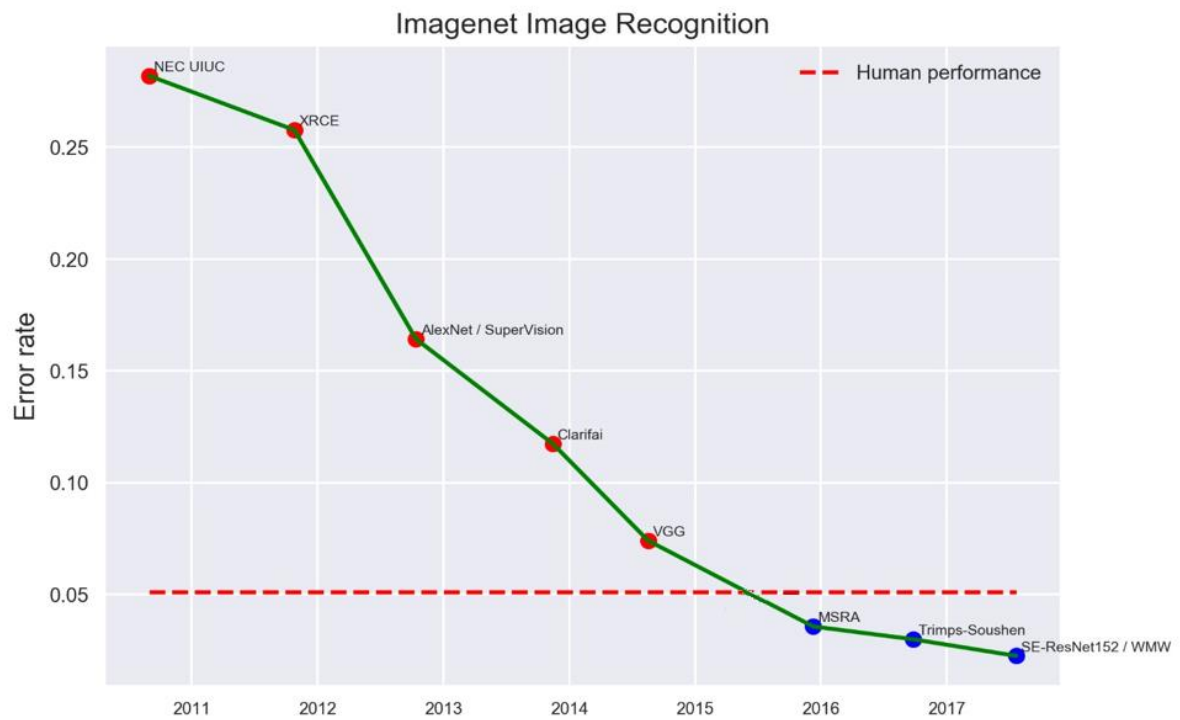
Vuonna 2012 AlexNet voitti ImageNet datachallengen merkittävällä marginaalilla käyttäen konvoluutionaalista neuroverkkoa [Krizhevsky *et al.* 2012]. Konvoluutionaaliset neuroverkot keksittiin jo vuonna 1980, ottaen mallia kissaeläinten visuaalisesta prosessointijärjestelmästä: tietyt neuronit ovat vastuussa tietyistä alueista [Fukushima 1980]. Kuitenkin vasta AlexNetin myötä konvoluutioverkot vakiinnuttivat asemansa kuvantunnistuksessa. Perinteisessä konenäössä piirteet irrotetaan manuaalisesti ja

tilannekohtaisesti, esimerkiksi kuvassa 3 jokainen pikseli on piirre. Syväoppivat konvoluutioverkot löytävät harjoitusesimerkkien avulla olennaiset piirteet automaattisesti.

Perinteiset neuroverkot koostuvat syötekerroksesta, piilokerroksista ja tuloskerroksesta, joissa kaikki neuronit ovat täysin toisiinsa yhteydessä. Konvoluutioverkot koostuvat täysin toisiinsa yhteydessä olevien kerrosten lisäksi konvoluutiokerroksista ja koontikerroksista. Konvoluutiokerrokset muodostetaan suodattimien avulla. Jos alkuperäisen kuvan koko on esimerkiksi $28 \times 28 \times 3$, voidaan siihen asettaa $5 \times 5 \times 3$ filtti, jonka arvoksi muodostuu 75-ulotteinen pistetulo + painoarvo. Suodatinta askeletaan haluttu pikselimäärä kerrallaan, kunnes koko kuva on käyty läpi. Kuva voidaan reunustaa "nollapikseleillä", jotta konvoluutiossa ei menetetä reunoilta dataa. Konvoluutiokerroksia asetetaan useita putkeen. Jokainen konvoluutiokerros kasvattaa kuvan syvyyttä, joten tarvitaan myös koontikerroksia. Koontikerroksia on erilaisia, mutta yleisin on maksimikoontikerros. Myös koontikerroksissa kuva käydään läpi askeleittain $x \times x$ kokoisilla suodattimilla. Suodatin päästää jokaisella askeleellaan vain alueensa suurimman arvon jatkoon ja samalla pienentää kuvan kokoa. Nykyään neuroverkot voivat olla jopa satoja kerroksia syviä [Huang *et al.* 2016].

Konvoluutioneuroverkkojen lisäksi toinen läpimurto syväoppimisessä on grafiikkaprosessorin käyttäminen koulutusprosessissa prosessorin sijaan. Grafiikkaprosessorin käyttö tekee koulutusprosessista noin 70 kertaa nopeampaa, lyhentäen viikkojen koulutusprosessit vain muutamaan tuntiin [Raina *et al.* 2009]. Grafiikkaprosessorien avulla saavutetaan myös sama laskentateho merkittävästi halvemmalla: Stanfordin tekoälylaboratorio saavutti 33 tuhannella dollarilla vastaavan laskentatehon kuin Googlen 5 miljoonan dollarin datakeskuksessa [Brown 2015].

Kuvassa 5 näkyy Imagenet kuvantunnistushaasteen voittajia virhemarginaaleineen. Ihmisen virhemarginaali on merkitty punaisella katkoviivalla - Konvoluutioverkkojen kehityksen myötä tekoäly pystyy nykyään luokittelemaan kuvia jo ihmistäkin tarkemmin.



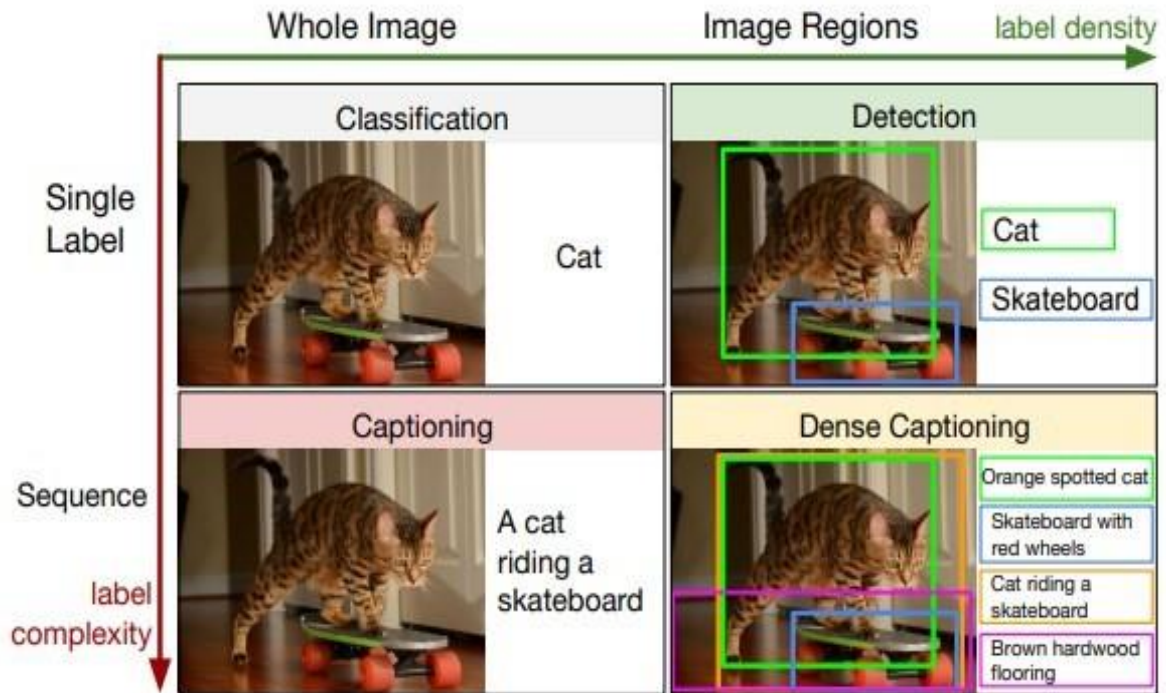
Kuva 5. Neuroverkot tunnistavat nykyään kuvia paremmin kuin ihmiset [Eckersley *et al.* 2017].

Kuvien syvempi ymmärrys vaatii kuitenkin paljon enemmän kuin kyvyn luokitella kuvia. Kuvat pitävät usein sisällään asioiden suhteita ja abstrakteja konsepteja. Sanonta “kuva kertoo enemmän kuin tuhat sanaa” kuvaa juuri tätä. Fiatin automainoksessa on auto metsässä ja ihmiset ihmettelevät oravaa kuin se olisi jokin peto. Kuvatekstinä on “We should all get out of the city from time to time”. Ihminen tietää, ettei orava ole vaarallinen ja kuva onkin hassu, koska mainoksen perhe näyttää pelkäävän sitä. Teksti viittaa myös siihen, että mainoksen auto on kykeneväinen kulkemaan muuallakin kuin kaupungissa.



Kuva 6. Fiatin [2019] automainos - “We should all get out of the city from time to time”

Tutkimus onkin lähivuosina siirtynyt kuvien syvempään ymmärtämiseen. Vuonna 2014 julkaistu COCO – Common Objects in Context pitää sisällään monimutkaisia arkitilanteita ja tyypillisiä asioita luonnollisessa kontekstissaan valmiiksi kuvailtuna [Yi *et al.* 2014]. Tämän avulla neuroverkkoja voidaan kouluttaa kuvailemaan kuvia laajemmassa kontekstissa yksityiskohtaisemmin. Johnson ja muut [2015] yhdistivät kappaleiden havaitsemisen ja kuvailun. Miksi kuvailla kuvaa vain yhdellä lauseella, kun kuvassa on paljon muutakin? Johnson ja muut [2015] pyrkivät luomaan kuvasta useita kuvauksia. Ensin kuva luokitellaan, jonka jälkeen havaitaan eri kappaleet kuvassa ja pyritään ymmärtämään niiden suhteita tuottaen useita kuvauksia luonnollisella kielellä. Tätä voidaan hyödyntää esimerkiksi kuvien käännteishaussa. Kuvien käännteishaussa hakuterminä on kuva itsessään ja hakukone pyrkii palauttamaan hakutuloksissa saman kuvan muista sivustoista, tai samankaltaisia kuvia. Kun kuvasta voidaan muodostaa useita luonnollisen kielen lauseita, voidaan tehdä entistä tarkempia kuvahakuja ja palauttaa tarkempia käännteishaun tuloksia.



Kuva 7: Dense captioning [Johnson *et al.* 2015]

Kuvien kuvailu kielellisellä ulostulolla toteutetaan purkamalla kuvan ominaisuudet konvotioaalaisella neuroverkolla ja syöttämällä tämä ulostulo rekursiiviseen neuroverkkoon, joita käytetään luonnollisen kielen tuottamiseen.

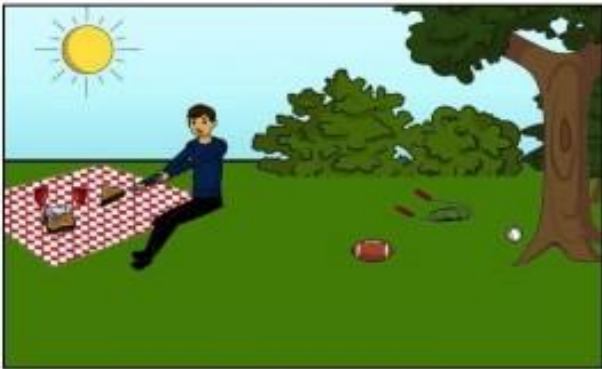
Kappaleiden tunnistus on jo sillä tasolla, että tarvitaan uusia testiasetelmia mittaamaan konenäön kehitystä matkalla kuvien täydelliseen ymmärtämiseen. Yksi näistä on COCO:n Visual Question Answering, jossa pyritään ymmärtämään kuvia vastaamalla kysymyksiin kuviin liittyen. Kuvassa 8 on esimerkkejä kisan kysymyksistä. Esimerkiksi odottaako kuvan henkilö seuraa? Ihmiselle kysymys on helppo: henkilö on piknikillä, joka on useimmiten seurallinen tapahtuma, joten henkilö todennäköisesti odottaa seuraa.



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



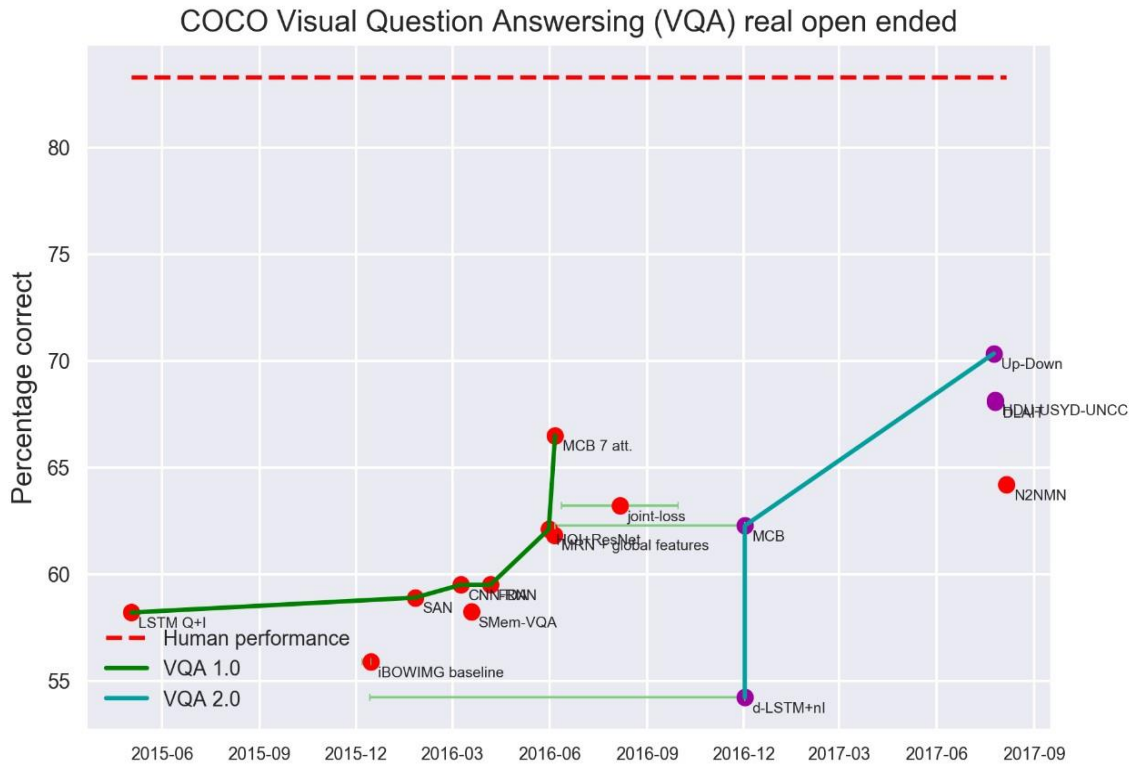
Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

Kuva 8. Visual Question Answering kysymyksiä. [Agrawal *et al.* 2015]

Toisin kuin kuvien luokittelutehtävissä, konenäkö ei ole vielä ihmisen tasolla (Kuva 9). Tehtävä on merkittävästi vaikeampi; tarvitaan kontekstuaalista ymmärrystä ja abstrakteja malleja kuvissa olevien kappaleiden suhteista.



Kuva 9. Visual Question Answering -tilastoja [Eckersley *et al.* 2017].

4 Yhteenveto

Visuaalinen sisältö on keskeinen väylä vaikuttaa kuluttajiin: kuvat ja videot dominoivat verkkosivustoja, applikaatioita ja sosiaalista mediaa. Yli 80 % internetin datasta on kuvia ja videoita [Cisco VNI 2018]. Internetin myötä mainontaa on kohdennettu keräämällä käyttäjäkohtaista dataa evästeiden, henkilökohtaisten tietojen, avulla. Lähivuosina yksityisyys on kuitenkin noussut pinnalle ja yhä useammat selailevat internettiä käyttäen evästeettömiä istuntoja, inkognito tiloja. Nykyään jo yli 60 % internetin käytöstä tapahtuu mobiililaitteiden kautta, mikä entisestään vaikeuttaa evästeiden käyttöä: appien evästeet eivät kommunikoi keskenään ja esimerkiksi Applen Safari-selain estää kokonaan kolmannen osapuolen evästeet [GumGum 2018, Sands 2015]. Myös lakisäädökset, kuten yleinen tietosuojasetus GDPR vaikeuttavat henkilökohtaisen datan keruuta, tehden mainonnan kohdentamisen selailutietojen pohjalta erittäin haastavaa.

Konenäkö luonnollisen kielen käsittelyn kanssa tarjoaa ratkaisun mainonnan kohdentamiseen: kun konenäkö pystyy ymmärtämään internetsivujen sisällön ja

kontekstin, markkinointimateriaali voidaan asettaa automaattisesti relevantteihin paikkoihin oikeaan kontekstiin. Manuaalisesti tämä on laajassa mittakaavassa mahdotonta, mutta jos konenäkö ymmärtää internetin visuaalisen median sisällön, helpottuu kontekstuaalinen kohdentaminen huomattavasti. Kontekstuaalisen mainonnan kohdentamisen on todettu olevan myös 73 % tehokkaampaa, kuin evästeiden pohjalta kohdistetun mainonnan [GumGum 2018].

Konenäön tehokas kuvien ja videoiden ymmärtäminen auttaa myös suojelemaan brändiä epätoivotuilta yhteyksiltä: pahimmissa tapauksissa brändi yhdistetään täysin ei-toivottuun materiaaliin, kuten L'Oréalin tapauksessa yrityksen mainokset pyörivät terroristijärjestöjen YouTube-videoilla [BBC 2017]. Kun konenäkö pystyy ymmärtämään verkkosivujen sisällön, voidaan mainokset asettaa haluttuun kontekstiin, brändille sopivaan ja turvalliseen ympäristöön.

Konenäön tutkimuksen alkaessa ei osattu aavistaa, kuinka vaikeaa konenäkö olisi täysin ratkaista. Vuosikymmeniä myöhemmin vuonna 2012 AlexNet muodosti konvoluutioverkot konenäön uudeksi paradigmaksi, mahdollistaen tien yleiskäyttöisempiin konenäön sovelluksiin kapea-alaisten tiettyyn tarkoitukseen tarkoin rakennettujen arkkitehtuurien sijaan. Konenäkö kykenee jo luokittelemaan objekteja paremmin kuin ihminen, muttei ole vielä saavuttanut ihmisen kykyä hahmottaa asioiden laajoja kokonaisuuksia ja monimutkaisia konteksteja. Konenäön hyödyntäminen markkinoinnissa on vasta aluillaan – vain 12 % yrityksistä käyttää konenäköä markkinoinnissaan [GumGum 2018]. Konenäön kehittyessä ihmisen tasolle visuaalisen datan prosessoinnissa ja ymmärtämisessä, tulee konenäkö varmasti näyttämään suurempaa roolia markkinoinnin apuvälineenä.

5 Viiteluettelo

Agrawal, Aishwarya, Lu, Antol, Mitchell, Zitnick and Batra. 2015. *VQA: Visual Question Answering*.

Anderlson. 1995. Checker Shadow Illusion.

<http://persci.mit.edu/gallery/checkersshadow> accessed on 30.5.2019.

BBC. Google apologizes after ads appear next to extremist content. 2017.

<https://www.bbc.com/news/business-39325916> accessed on 30.5.2019.

Brown. 2015. GPU ACCELERATED DEEP LEARNING WITH CUDNN.

<https://www.slideshare.net/NVIDIA/gpuaccelerated-deep-learning-for-cudnn-v2>
accessed on 19.4.2019.

Chen, 2016 *Handbook of Pattern Recognition and Computer Vision (5th Edition)*. World Scientific.

Cisco, Visual Networking Index. 2018.

<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html> accessed on 17.4.2019.

Deng, Dong, Socher, L.-J. Li, K. Li and L. Fei-Fei. 2009. ImageNet: A LargeScale Hierarchical Image Database. In *IEEE Computer Vision and Pattern Recognition (CVPR), 2009*.

Domo. 2018. Data Never Sleeps 6.0.

<https://web-assets.domo.com/blog/wp-content/uploads/2018/06/18-domo-datanever-sleeps-6.png> accessed on 17.4.2019.

Eckersley and Nasser. 2017. EFF AI Progress Measurement Project, <https://eff.org/ai/metrics>, accessed on 17.4.2019.

Fiat. 2019. We should all get out of the city from time to time. www.fiat.com accessed on 30.5.2019.

Fukushima. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics* 36; 193–202.

- Grobstein. 2003. Getting It Less Wrong, the Brain's Way: Science, Pragmatism, and Multiplism. In *Ritivoi, A.D. (ed) Interpretation and Its Objects: Studies in the Philosophy of Michael Krausz*, (pp. 153-166). New York.
- GumGum. 2018. Seen & Unseen: How Marketers Use Computer Vision <https://www.gumgum.com/guides/seen-unseen> accessed on 30.5.2019.
- Huang, Gao, Sun, Liu, and Sedra. 2016. Deep Networks with Stochastic Depth. In *European conference on computer vision* (pp. 646-661). Springer, Cham.
- Huang. 1996. Computer Vision: Evolution and Promise. In *19th CERN School of Computing* (pp. 21-25). Geneve.
- Johnson, Karpath and Fei-Fei. 2016. *DenseCap: Fully Convolutional Localization Networks for Dense Captioning*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4565-4574).
- Krizhevsky, Sutskever and Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Neural Information Processing Systems 25 (NIPS 2012)*.
- Krulwich. 2012. Which Is Greater, The Number Of Sand Grains On Earth Or Stars In The Sky? <https://www.npr.org/sections/krulwich/2012/09/17/161096233/which-is-greater-the-number-of-sand-grains-onearth-or-stars-in-the-sky?t=1555486111924> accessed on 17.4.2019.
- Marr. 2010 *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. The Mit Press.
- Nielsen. 2015. *Neural Networks and Deep Learning*. Determination Press. Vol. 25.
- Papert. 1966. *The Summer Vision Project*. MIT Artificial Intelligence Group, Vision Memo No. 100, 1-6.
- Pinto, Cox and DiCarlo. 2008 *Why is Real-World Visual Object Recognition Hard?* PLoS computational biology, 4(2);1-27.
- Potter, Wyble, Haggmann and McCourt. 2014. Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophys* 2014, 76(2): 270-279.
- Raina, Rajat, Madhavan, Anand and Andrew. 2009. Large-Scale Deep Unsupervised learning using graphics processors. In: *Proceedings of the 26th International Conference On Machine Learning, (ICML), 2009*.
- Roberts. 1963 *Machine Perception of Three-Dimensional Solids* MIT Lincoln Laboratory Technical Report No 315.

Sands. 2015. How The Cookies Crumble In A Mobile-First World
<https://martechtoday.com/cookies-crumble-mobile-first-world-154114> accessed on
30.5.2019.

Serendipstudio. 1994. Ballshadow
<https://serendipstudio.org/bb/ballshadow/> accessed on 30.5.2019.

Yi, Maire, Belongie, Bourdev, Girshick, Hays, Perona, Ramanan and Zitnick. 2018.
Microsoft COCO: Common Objects in Context.