Juho Hermunen

Kasvojentunnistusohjelmien ongelmia ja niiden ratkaisuja käyttäen generatiivisia kilpailevia neuroverkkoja

Tietotekniikan kandidaatin tutkielma

21. toukokuuta 2021

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Juho Hermunen

Yhteystiedot: jujosahe@student.jyu.fi

Ohjaaja: Leevi Annala

Työn nimi: Kasvojentunnistusohjelmien ongelmia ja niiden ratkaisuja käyttäen generatiivisia kilpailevia neuroverkkoja

Title in English: Issues in facial recognition and solutions using generative adversial networks

Työ: Kandidaatin tutkielma

Opintosuunta: Kaikki opintosuunnat

Sivumäärä: 20+0

Tiivistelmä: Tässä kandidaatintutkielmassa selvitän neuroverkkoihin ja kasvojen tunnistukseen liittyviä ongelmia ja generatiivisten kilpailevien neuroverkkojen niihin tarjoamia ratkaisuja. Lisäksi tarkastelen miten teknologian kehitys voisi ratkaista joitain ongelmia ja mitä odotan tulevaisuudelta. Kasvojentunnistusohjelmat ovat yleistyneet monissa eri käyttökohteissa, mutta niiden tarkkuus ei ole halutulla tasolla. Tässä tutkimuksessa selvitän monia eri ratkaisuja ongelmiin, joista lupaavimpana koen 3D generatiivisen kilpailevan neuroverkon kyvyn luoda uutta aineistoa tehokkaasti.

Avainsanat: neuroverkot, kasvojentunnistus, kandidaatintutkielmat

Abstract: In this bachelor theses I will be going through issues related to neural networks and facial recognition while going through some solutions offered by generative adversarial neural networks. I will also be investigating how the development of technology might be able to solve some of these problems and what do I expect from the future of this technology. Facial recognition programs have become more and more common in different uses but the accuracy of them leaves much to be desired. In this paper I will be going through solutions for different problems of which I find 3D generative adversarial neural network to be the most promising with its ability to efficiently generate new data.

Keywords: Neuralnetwork,Facial recodnition,Bachelor's Theses

Kuviot

Sisällys

1 Johdanto

Mikä on syy, kun iPhonen lukko ei avaudu? Miksi massavalvonta askarruttaa ihmisiä ja miten lähellä sen mahdollista toteutusta ollaan? Kasvojentunnistus on yhä enemmän uutisissa, sillä yksityisyyden suojaan liittyvät kysymykset ovat nousseet esille erilaisissa tietovuodoissa. Kasvojentunnistusohjelmat ovat käytössä monissa eri käyttötarkoituksissa ja niitä voi joutua kohtaamaan tietämättään monissa eri tilanteissa. Kasvojentunnistusohjelmat eivät ole kuitenkaan täydellisiä, sillä niihin liittyy monia erilaisia ongelmia. Tämän tutkimuksen tarkoituksena on tulevaisuuden kartoittamisen lisäksi saada jonkinlainen kuva siitä, mitä kasvojentunnistusteknologia on tällä hetkellä ja ehdottaa mahdollisia ratkaisuja joihinkin näistä ongelmista.

Tässä tekstissä selitän lyhyesti mitä neuroverkot ovat, mitä konvoluutioverkot ovat ja mitä ongelmia ja ratkaisuja näihin löytyy kirjallisuudesta. En aio ottaa kantaa kasvojentunnistusohjelmien laillisuuteen tai niiden etiikkaan, sillä se monimutkaistaa keskustelua huomattavasti.

Aluksi selvitän mikä on neuroni ja mitä se sisältää, samoin sen ensimmäiset keinotekoiset toteutukset ja niiden heikkoudet. Käsittelen myös lyhyesti konvoluutioverkkojen perustoimintaperiaatteita, mitä sen kerrokset tekevät ja miten ne vaikuttavat lopputulokseen. Selvitän generatiivisen kilpailullisen neuroverkon toimintaa ja rakennetta, yleisiä ongelmia ja niihin olevia ratkaisuja. Huomion tekstissä saa myös se, mitä spesifejä ongelmia muodostuu kasvojentunnistusaplikaatiossa. Kappaleessa 3 selvitän, miten generatiivisia kilpailullisia neuroverkkoja voidaan käyttää ratkaisemaan joitain näistä ongelmista täydentämällä puuttuvaa dataa ja luomalla synteettistä dataa. Viimeiseksi pohdin mitä mahdollisuuksia käännetty generatiivinen kilpailullinen neuroverkko antaa yhdistettynä kappaleessa 3.2 esitettyyn 3D generatiiviseen kilpailulliseen neuroverkkoon. Tarkastelen myös mitä 3D generatiivinen kilpailullinen neuroverkko mahdollistaa, jos se voitaisiin saada sen teoreettiselle maksimitasolle.

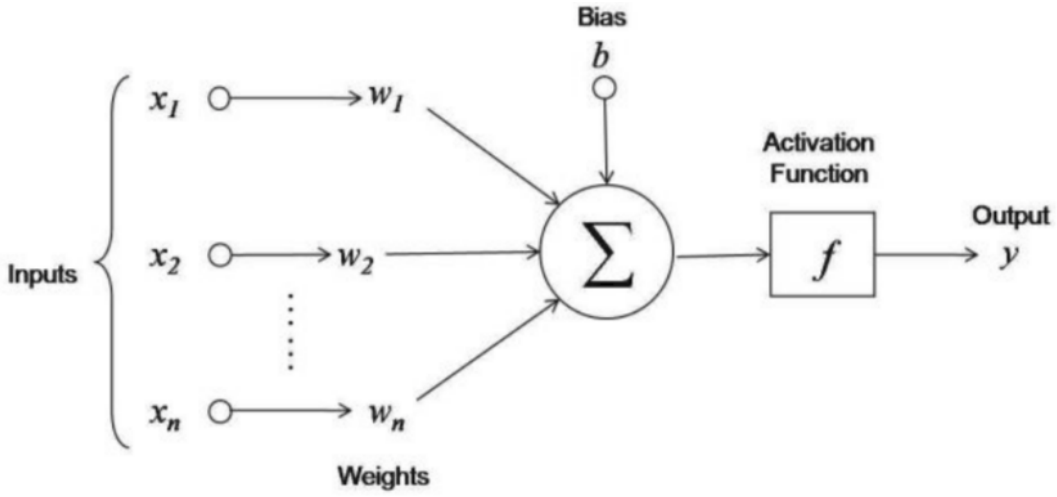
2 Neuroverkot

Neuroverkko on termi, joka on lähtöisin biologiasta ja sillä viitataan eliöiden keskushermostoihin(Hopfield 1982). Ne sisältävät neuroneja, joista nimi on lähtöisin. Tietotekniikassa neuroverkoilla tarkoitetaan algoritmeja, joilla tietokone pyrkii mukailemaan keskushermoston toimintaa. Tämä on hyödyllistä, kun tietokonetta tarvitaan ratkaisemaan erilaisia kategorisointiongelmia, näin tehden niistä tehokkaita konenäköalgoritmejä(Goodfellow, Bengio ja Courville 2016).

Ideaalitoteutuksessa neuroverkolle rakennettaisiin sille oma spesifisti suunniteltu laite, esimerkiksi kustomoitu mikropiiri(Lacey, Taylor ja Areibi 2016). Tämä mahdollistaisi Laceyn mukaan suurien tietomäärien läpikäymisen neuronien käsitellessä tietoa itsenäisesti, toisin kuin tavallisessa tietokonearkkitehtuurissa, missä CPU käsittelee yhden operaation kerrallaan. Tehtävän toteuttamista auttaa myös se, että neuronit sisältävät hieman lyhytaikaista muistia ja niiden väliset yhteydet sisältävät pitkäaikaisen muistin painokertoimina. Tätä ei kuitenkaan ole mahdollista toteuttaa jokaiselle neuroverkolle. Joten nykyisellään neuroverkot, jotka toimivat tietokoneilla, joutuvat turvautumaan perinteisiin CPU- operaatioihin, koska niiden GPU- laskentateho ei ole riittävä(Lacey, Taylor ja Areibi 2016). Neuroverkko sisältää syötekerroksen vähintään yhden piilotetun kerroksen ja tulostekerroksen. Nämä on tavallisesti liitetty toisiinsa niin, että jokaisen neuronin tuloste on seuraavan neuronin syöte pois lukien tulostekerros (Jain, Mao ja Mohiuddin 1996).

Neuroni sisältää tavallisesti syötteen, painokertoimen, painotuksen ja tulosteen. Painokertoimet ovat osa neuroverkkoa, joka muuttaa saatua syötettä oppivassa neuroverkossa johonkin suuntaan saavuttaakseen halutun lopputuloksen. Painotus on kuin vakiotermi funktiossa eli se ohjaa toimintaa tiettyyn suuntaan. Niillä on myös oltava jonkinlainen aktivaatiofuntio joka ei muutu ja jota käyttäen saadaan neuronin tuloste. Neuroni toimii saamalla jonkin syötteen, johon se soveltaa painokerrointa ja painotusta muokaten syötettä. Sen jälkeen se käyttää aktivaatiofuntiota muokattuun syötteeseen, näin ollen tuottaen tulosteen.

Ensimmäisiä keinotekoisia neuronitoteutuksia oli perseptroni, joka kehitettiin Amerikassa 1950-luvulla (Rosenblatt 1958). Se on yksinkertainen metodi, jossa perseptronineuroni op-



Kuvio 1. Neuroni. (De Oliveira ym. 2017)

pii asioita nimensä mukaisesti perseptronialgoritmia käyttäen. Tämä oli ensiaskel kuvien tunnistuksen saralla. Ongelmaksi nousi se, että painokertoimien muuttaminen miljoonille tarvittaville neuroneille ei ollut perinteisellä perseptronialgoritmilla tehokasta.

Tähän ongelmaan ratkaisuksi tuli monikerrosperseptroni (Goodfellow, Bengio ja Courville 2016). Se käyttää takaisinvirtausalgoritmia (backpropagation) opettamiseen tavallisen perseptronialgoritmin sijaan (Bello 1992). Tämä mahdollistaa opettamisen antamalla verkolle dataa, jonka mukaan se päivittää painoja ja jatkaa tätä jokaisen kerroksen kohdalla. Tämä eroaa perinteisestä algoritmista, koska se lähestyy verkkoa sen lopusta näin ollen vähentäen turhan laskemisen määrää ja gradienttia voidaan laskea kokonaisuudessaan paloina. Monikerrosperseptroni on eteenvirtaus (feed-forward) verkko eli sen neuronit eivät muodosta kehää vaan kaikki syötteet johtavat johonkin tulosteeseen.

2.1 Konvoluutioverkko

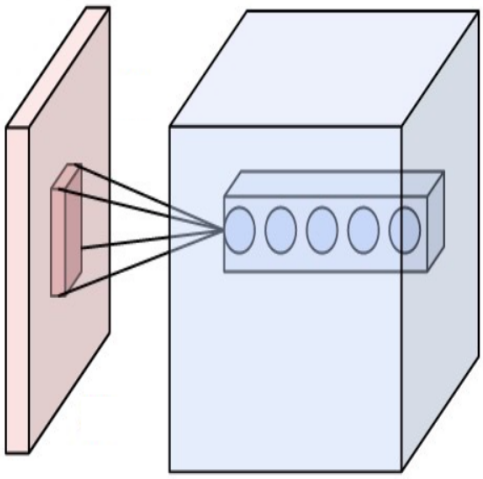
Konvoluutioverkko, josta tulen käyttämään lyhennettä CNN (Convolutional Neural Network), on monikerroksinen neuroverkko, jossa on vähintään yksi konvoluutio kerros. CNN historia alkaa 1980-luvulla japanilaisen tietoteknikon Kunihiko Fukushiman kehittäessä neokognition (Fukushima ja Miyake 1982) ja tätä voidaan pitää yksinkertaisena konvoluutio neuroverkkona. Siinä oli ensimmäistä kertaa konvoluutiokerros ja kokoamiskerros (downsampling

layer). Se mahdollisti sen toiminnan paremmin erilaisissa tehtävissä ja se toimi myös alkupisteenä tutkimukselle konvoluution käyttöön kuvankäsittelyssä.

CNN on lähes aina osakonnektiivinen verkko eli sen yhden kerroksen neuronit eivät ole yhteydessä kaikkiin seuraavan kerroksen neuroneihin. Tämä johtaa yleisesti ongelmiin ylisovittamisen kanssa, minkä takia verkosta voi tulla toimiva vain tietylle datalle ja sen toiminta heikkenee muille datoille (Lawrence, Giles ja Tsoi 1997). Tämä on muodostunut ongelmaksi lähes kaikissa monikerrosperseptroniverkoissa ja siihen on löydetty erilaisia ratkaisuja. Tätä ongelmaa ratkaistaan regularisoimalla aineistoa erilaisilla algoritmeilla (Goodfellow, Bengio ja Courville 2016). Ongelma on ratkaistu useimmiten konvoluutioverkossa kokoamisella (pooling), joka tapahtuu kokoamiskerroksessa koska parametrien väheneminen vähentää regularisoinin tarvetta. CNN aktivaatiofunktiona käytetään lähes aina ReLU (Rectified Linear Unit) (Nwankpa ym. 2018), joka korvaa negatiiviset aktivaation arvot nollalla ja näin ollen vähentäen aktivaatioiden määrää ja siten vähentää laskettavuutta.

Konvoluutionkerros on kerros, jossa osa neuroneista on havainnoimassa edellisen kerroksen osaa ja minkä painotusvektoreista käytetään termiä filter. Tämä kehitettiin ratkaisemaan laskennallisuusongelmia, sillä perinteisessä monikerrosperseptronissa jokainen neutroni sisältää oman painonsa. Koska neuronit ovat kokonaiskonnektiivisia, neuronien painojen määrä kasvaa exponenttiaalisesti neuronien määrän kasvaessa. Tämä arkkitehtuuri ei ota myöskään huomioon luonnollisissa kuvissa olevaa luonnollista paikallisuutta (princible of locality), eli pikselit vaikuttavat toisiinsa. Sillä jos pikseli, joka on kuvassa korvasta ja esiintyy havainnointikentässä toisen pikselin lähellä, mikä on edellä mainitussa kuvassa, nousee todennäköisyys havainnointikentässä olevan asian esittävän korvaa paljon enemmän kuin tapaus, jossa nämä samat pikselit olisivat eri havainnointikentissä. Perinteisessä tiheässä neuroverkossa kaikki pikselit ovat samanarvoisia ja tällainen jaottelu on mahdotonta.

Konvoluutiokerros ratkaisee tätä ongelmaa asettaen neuronit havainnoimaan rajattua aluetta kuvassa. Neuronien havainnoidessa vain tiettyä kohtaa edellisessä kerroksessa samalla kerroksella olevat neuronit katsovat samaan aikaan eri paikkaa kerroksella, mutta kloonatuilla painoilla. Näin niiden käyttö on paljon laskenta tehokkaampaa koska yhteyksiä on vähemmän. Tämä auttaa myös tunnistamaan klustereita kuvissa, jotka vastaavat jotain asiaa eri kohdissa kuvaa ja mahdollisesti muuttuneissa olosuhteissa (Albawi, Mohammed ja Al-Zawi

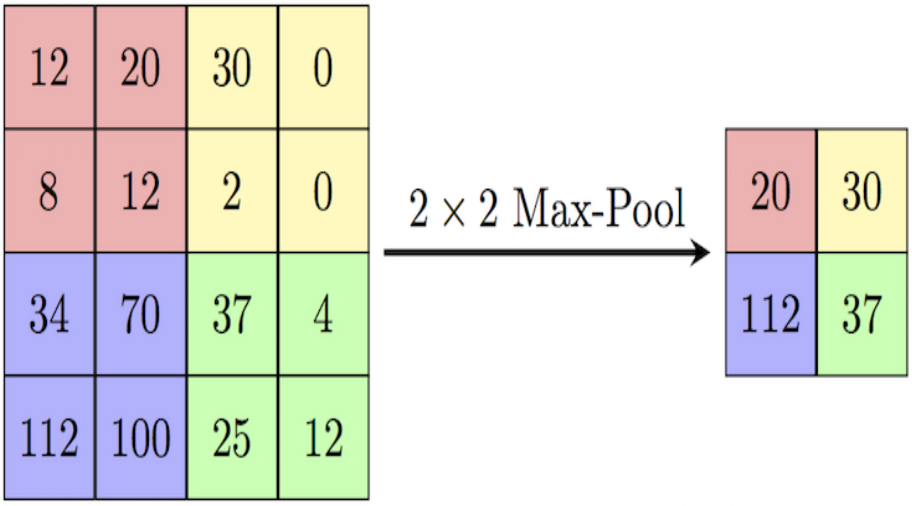


Kuvio 2. Konvoluutiokerroksen toiminta. (Albawi, Mohammed ja Al-Zawi 2017)

2017).

Tätä operaatiota, jossa fillteriä käytetään koko kuvaan osittain, kutsutaan konvoluutioksi ja se tuottaa havainnointikartan (feature mapin)(Albawi, Mohammed ja Al-Zawi 2017). Siihen tallennetaan aktivaatioiden määrä tällä kuvan osalla. Seuraavaksi seuraava kerros käy taas koko kuvan läpi omilla kloonatuilla painoillaan luoden näin uuden havainnointikartan. Tämä mahdollistaa tunnistuksen koko kuvan alueelta, vaikka filtteri olisi opetettu datalla missä korva on aina kuvan vasemmassa laidassa, niin konvoluutio löytää sen myös kuvan oikeasta laidasta.

Kokoamiskerros on kerros missä neuronit havainnoivat osaa havainnointikartasta. Ne laskevat kerroksen neuronien aktivointien määrää ja tällä voidaan valita mikä alue on eniten aktivoitu ja näin ollen päätellä mitä asiaa jokin arvo eniten kuvaa (Albawi, Mohammed ja Al-Zawi 2017). Tämä perustuu yleistykseen missä kerros hylkää osan havainnoista havaintokartassa. Nykyään kokoamiskerros käyttää useimmiten maksimi aktivaatioiden määrää (Max pooling) yleistääkseen eikä aktivaatioiden keskiarvoa. Tämä auttaa vähentämään tarvittavaa laskentatehoa, sillä kokoamiskerros toimii konvoluutiokerroksen tulostekerroksena. Kokoamiskerroksen tulostuskerroksessa se luo pelkistetyn kuvan havaintokartasta ja sen jälkeen se antaa saadun tulosteen seuraavalle kerrokselle (Albawi, Mohammed ja Al-Zawi 2017).



Kuvio 3. Max poolingin toiminta. (Albawi, Mohammed ja Al-Zawi 2017)

2.2 Generatiivinen kilpaileva verkko

Generatiivinen kilpaileva verkko, josta tulen käyttämään lyhennettä GAN (Generative Adversial Network), on konsepti, jonka kehitti Goodfellow ym. (2014). Sen juuret voidaan katsoa olevan signaalin tulkinnassa, koska sekin käyttää puuttuvan datan täydentämistä. Siinä on kaksi verkkoa: tuottava- ja erottavaverkko ja ne opetetaan opetusdatalla joko luomaan kuvia tai erottamaan mikä kuva kuuluu opetusdataan ja mikä ei. Tämä johtaa tuloksesta riippuen jommankumman verkon painojen muutokseen missä se pyrkii muuttamaan toimintaansa niin, että se parantaa suoriutumistaan seuraavassa sarjassa kuvia. Tämä on nollasummanpeli eli toisen verkon on aina voitettava ja toisen on aina hävittävä. Tämä johtaa teoriassa olettaen äärettömän määrän muistia tilanteeseen, missä verkko ei voi enää parantua koska generoitu data vastaa oikeaa dataa. Tämä on mahdollinen tapa luoda uutta dataa, jolla voidaan toteuttaa erilaisia mallien opettamisfunktioita ilman tarvetta kerätä massiivista määrää dataa valmiiksi. Sen on näytetty toimivan äärimmäisen tehokkaasti kuvien puuttuvan datan täydentämisessä, joten uskon sen myös toimivan muissa jossain määrin hierarkkisissa datan tuottotehtävissä (Isola ym. 2017).

Tuottavan verkon tehtävänä on luoda uusia kuvia. Näitä uusia kuvia näytetään sen jälkeen sekoitettuna opetusdataan erottelevalle verkolle ja se pyrkii erottamaan ne toisistaan. Se on tyypillisesti dekonvoluutioverkko, joka toimii käänteisesti konvoluutioneuroverkkoon nähden

eli se luo puuttuvaa dataa pienemmästä datamäärästä dekonvoluutiolla. Tämä on mahdollista levityksellä (upsampling). Siinä pieni määrä dataa asetetaan isommalle alueelle käyttämällä erilaisia algoritmeja, kuten osalineaarinen interpolointi (Bi-Linear Interpolation). Sillä saadaan rakennettua perusrakenteet kuvan luomiselle ja jota painoja siirtämällä voidaan muokata tulosten mukaan.

GANin ongelmia ovat ne, että niiden testaaminen toisella GANilla on hankalaa, sillä se vaatisi kahden verkon oppimisen samassa tahdissa ja tätä on hyvin vaikea hallita ohjaamattomalla oppimisella (Arjovsky ja Bottou 2017). Tämä on yksinkertaisesti seuraus verkon rakenteesta ja sen antamista hyödyistä, mutta tähän on ehdotettu ratkaisuksi SGAN (Supervised GAN). Siinä opetetaan monia pareja itsenäisesti samalla aineistolla ja koulutetaan niitä käyttäen yksi valvova GAN (Chavdarova ja Fleuret 2018).

Toiseksi ongelmana voi olla tutkittavuus: jos vaikka verkko tuottaa viisi samanväristä ja -kuviollista kuvaa, jotka näyttävät ihmissilmään hyvin samanlaisilta, voivat ne olla verkon mukaan eri kuvia tai se voi vain ylikompensoida diskriminoivaa verkkoa kohtaan jättäen itsensä silmukkaan (Wang ym. 2017). Tähänkin on kuitenkin jonkinlaisena ratkaisuna Wasserstein GAN, joka osittain ratkaisee testattavuuden ja tulkittavuuden ongelmat (Arjovsky, Chintala ja Bottou 2017).

2.3 Ongelmia kasvojen tunnistuksessa

Ensimmäinen ongelma, mihin kasvojentunnistusohjelmat törmäävät, on datan laadun heikkous. Sillä kaikkea dataa ei ole mahdollista saada ihanteellisissa olosuhteissa, varsinkin jos on kyse luonnollisista kuvista. Tämä on suurin ongelma, sillä on tunnettua, että huonommin optimoitu algoritmi voi antaa parempia tuloksia kuin paremmin optimoitu algoritmi, jos sen saaman opetusdatan laatu on parempaa (Martinez ja Valstar 2016). Ongelmia muodostuu myös edellä mainituista laadun heikkouksista lisäten luokansisäistä varianssia (intra-class variance). Näitä ovat valaistusolosuhteet, erilaiset pään asennot, kameran laatu, linssin vääristymät ja muu kohina.

Seuraava ongelma on leimojen yhdistäminen dataan. Koska neuroverkkojen opettamiseen tarvittavan datan määrä on hyvin suuri, on sen manuaalinen leimaaminen hankalaa tai li-

ki mahdotonta ja silloin on leimaaminen toteutettava jollakin automaattisella menetelmällä. Yleisesti tämä tarkoittaa jonkinlaisen ohjelman käyttämistä. Tämä ohjelma pystyy leimaamaan dataa itsenäisesti tai vain pienellä ohjauksella.

Ongelmia muodostuu myös erilaisista leimaamiskriteereistä. Tällä tarkoitetaan erilaisten kasvokuviin liitettävien leimojen subjektiivisuutta. Tämä on erityisen suuri ongelma ilmeiden tunnistamisessa, sillä ihmisten ilmeiden subjektiivisuus on hyvinkin suuri.

Viimeinen yleinen ongelma on kasvojentunnistusohjelman painottuminen tietylle datalle (Martinez ja Valstar 2016). Koska lähes kaikkia kasvojentunnistusohjelmia testataan yleisesti saatavilla olevia massiivisia datasettejä vastaan ja jokainen näistä dataseteistä sisältää jossain määrin painotusta eri asioissa, on mahdollista, että kasvojentunnistusohjelmaa ohjataan tiedostamatta toimimaan hyvin tälle datasetille laiminlyöden sen yleistyksen kaikelle datalle.

3 Ratkaisuja

Tässä kappaleessa käsittelen erilaisia tutkimuksia, jotka voivat ratkaista kappaleen 2.3 mainituista ongelmista. Ei ole kuitenkaan olemassa ”hopealuotia”, jolla saisimme ratkaistua kaikki ongelmat samaan aikaan. Joten jokaiseen ongelmaan on etsittävä erillisiä ratkaisuja ja niitä on mahdollisuuksien mukaan ydisteltävä. Tämän lisäksi, koska GAN on käsitteenä hyvin uusi, on siihen liittyvä tutkimus jossain määrin alkutekijöissään ja sen kehitys on huimaa.

Seuraavaksi käsittelen erilaisia GAN variantteja, joiden käyttötarkoituksia pohdin kappaleen viimeisessä luvussa.

3.1 Kuvien parantaminen GANilla

Kuvien laatu voi olla huono monesta eri syystä: erilaiset valaistusolosuhteet, kameran linssin heikko laatu tai vain kameran vähäinen pixelimäärä. Tätä ongelmaa voidaan lieventää täydentämällä kuvia käyttämällä cGAN (conditional) (Isola ym. 2017) kuvien parantamiseen (Image-to-Image Translation).

Isolan tutkimuksessa he kokeilivat myös pelkän L1 regularisoinnin käyttöä ja L1 regularisoinnin sekä cGAN yhdistelmää. L1 regularisoinnilla savutetaan kohtuullisia tuloksia, mutta se johtaa kuvien generoinnissa sumeisiin kuviin. Se ei myöskään täydennä kuvia erilaisilla yksityiskohdilla, jotka eivät ole sille annetussa syöttökartassa (input label map), toisin kuin cGAN, joka täydentää asioita kuvaan, jotka eivät olleet syötekartassa lisäten kuvan tarkkuutta tehden rajoista (outline) selkeämpiä. Pelkän cGAN käyttö antaa parempia tuloksia kuin L1, mutta kaikista parhaimpia kuvia saatiin yhdistämällä molemmat tekniikat (Isola ym. 2017).

Tästä hyvänä esimerkkinä toimii (Yang, Zhang ja Yin 2018) tutkimus ilmeiden generoinnissa. Tässä tutkimuksessa cGAN oli juuri edellisen kappaleen mukaan valjastettu täydentämään puuttuvaa dataa annetun syötteen perusteella. Tässä tutkimuksessa cGAN oli opetettu tuottamaan kuusi perusilmettä henkilön kuvasta ja näitä verrattiin julkisiin tietokantoihin, kuten OULU-Casia. Tarkoituksena oli tuottaa tapa, jolla voidaan tuottaa uusia ilmeitä niin

Kuvio 4. Kuva eri tavoilla opetettujen verkkojen toiminnasta. (Isola ym. 2017)

että malli mukautuisi eri henkilöille (model is identity-adaptive).

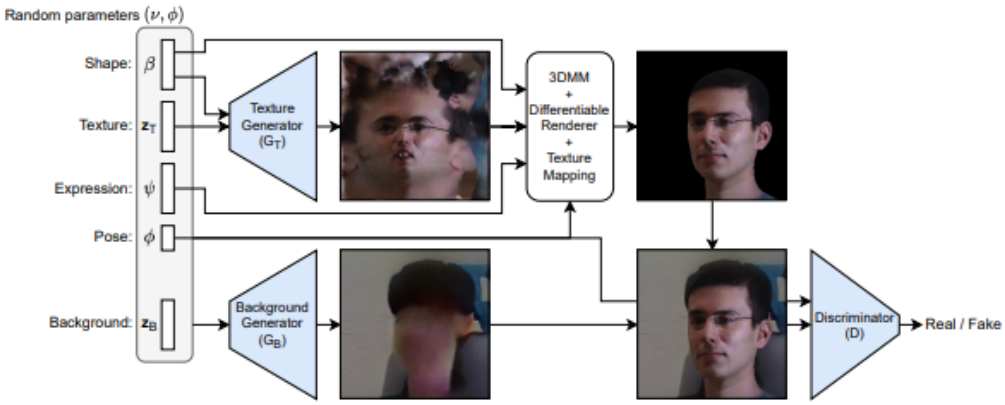
Tämä voisi ratkaista datasettien heikkolaatuisuutta suoraan korjaamalla datasettien puutteita parantamalla niissä olevien kuvien laatua. Pelkästään kuvien levitys (upsampling) auttaisi konvoluutioverkkoja tekemään tarkemman havaintokartan.

3.2 Synteettisen datan luominen GANilla

Synteettistä dataa voidaan luoda monilla eri tavoilla, mutta tässä kappaleessa keskityn siihen, miten sitä voidaan luoda 3D-GANilla ja miten synteettinen data vaikuttaa neuroverkoon jos sitä käytetään opettamisvaiheessa

3D-GAN eroaa tavallisesta GANista siinä, että toisin kuin perinteisessä GAN- kuvassa, jossa kolmeulotteisuutta pystytään vain jäljittelemään 2D-elementtien päällekkäin asettelulla. 3D-GAN koostuu kolmesta verkosta: tavallisen erottelijaverkon lisäksi on verkko, joka luo kolmeulotteisen kuvan ja kolmas verkko, joka luo taustan (Marriott, Romdhani ja Chen 2020). 3D-GAN toimii ottamalla pohjakuvasta kasvonpiirteitä(feature) ja jotka se asettaa 3D-mallinnusohjelmaan, joka oli tässä tutkimuksessa FLAME (Faces Learned with an Articulated Model and Expressions). Se luo manipuloitavan 3D-objektin, josta se voi poimia erilaisia ilmeitä ja kuvakulmia. Sitten se liittää tämän kuvan toisen tuottajaverkon luomaan taustaan, jonka jälkeen siitä otettu kuva annetaan erottelijaverkolle ja sen tulosten mukaan painoja muutetaan.

3D-GAN voidaan kouluttaa samalla tavalla kuin tavallinen GAN käyttäen perinteiseen GANiin sovellettavia metodeja, kuten Wasserstein. Uniikkina kontrollimetodina toimii kuvakulmien vaihtaminen jokaisella kierroksella, minkä avulla pyritään vähentämään ylisovittu-



Kuvio 5. Kuva 3D-GANin toiminnasta. (Marriott, Romdhani ja Chen 2020)

mista. Ongelmia muodostuu kuitenkin tiettyjen piirteiden, jotka ihmismieli liittää kasvoihin, mallintamisessa. Näitä ovat esimerkiksi karvat, hampaat ja silmälasit, sillä niitä ei ole liitetty FLAME- malliin ja ne esiintyvät siinä vain tekstuureina ja ne voivat aiheuttaa ongelmia eri kulmista. Nykyisellään tämän tutkimuksen 3D-GAN oppii valaistusolosuhteet osaksi pohjakuvaa. Tämä johtaa vääristymiin eri kuvakulmista, sillä ne tulevat osaksi verkon rakentamia kasvoja.

3D-GANia voidaan hyödyntää synteettisen datan luomisessa, näin sitä pystytään käyttämään kasvojentunnistuksen parantamisessa (Kortylewski ym. 2018). Tutkimuksessa todetaan kuitenkin, että täysin synteettisen datan käyttö saavutti heikompia tuloksia kuin pelkän tavallisen datan kanssa. Tutkijat kuitenkin tulivat johtopäätökseen, että yhdistämällä synteettistä ja tosi maailman dataa saatiin aikaan parempia tuloksia kuin kummallakaan datatyypillä yksistään ja synteettisen datan käyttö voisi vähentää tosi maailman datan tarvittavaa määrää.

3.3 Pohdinta

(Kortylewski ym. 2018) tutkimuksessa mainittiin ongelmaksi, että kasvojen generointiohjelmien heikkoudet olivat yksi syy synteettisen datan heikkouteen verrattuna tosimaailman dataan. 3D-GAN pystyisi ehkä ratkaisemaan näitä ongelmia, sillä sen kyky generoida realistista dataa on (Marriott, Romdhani ja Chen 2020) tutkimuksessa todettu hyväksi. Tämä mahdollistaisi ehkä täysin synteettisen datan käyttämisen ilman huomattavaa vähenemistä kasvojen

tunnistuksen laadussa. Tämä voisi ratkaista ongelmia puutteellisten datasettien osalta, kuten vaikka tunnettuja puutteita eri ihonvärin omaavien henkilöiden määrässä ja näin ollen mahdollistaen ”globaalin” datasetin.

Tästä voisi teoriassa seurata Goodfellow ym. mainitsema tuotetun datan vastaavuus, jolloin saisimme luotua täydellistä dataa lisää ja sitä kautta rakennettua täydellisen datasetin, jolla voisimme ohittaa kaikki ylisovittamisongelmat. Käytännössä tämä ei ole tällä hetkellä mahdollista oman tietämykseni mukaan, sillä nykyinen laskentateho tietokoneissa on rajoittava tekijä. Uskon jonkinlaisen lähes täydellisen datasetin kuitenkin olevan tulevaisuudessa mahdollinen ratkaisu data-ongelmiin, joita olen tässä tutkimuksessa maininnut. Täydellinen datasetti ei ole tällä hetkellä tehokas ratkaisu verrattuna algoritmien parantamiseen.

Toinen mahdollisuus on käyttää 3D-GANiin (Creswell ja Bharath 2018) käännettyä (inverted) GANia, jolla voidaan yrittää löytää erilaisia parametrejä kuvista. Tämä voisi ratkaista osan leimaamisongelmista, jos GANista saataisiin järkeviä parametrejä mitkä voitaisiin leimata ja millä voisi opettaa muita neuroverkkoja. Tämä voisi ratkaista joitain leimaongelmia mitä massiivinen määrä dataa sisältää.

4 Yhteenveto

Tässä tutkimuksessa pohdittiin ongelmia, joita neuroverkkopohjaiset kasvojentunnistusohjelmat kohtaavat erilaisissa tilanteissa ja miten niitä voitiin ratkaista käyttäen GANeja. Sen lisäksi esiteltiin yleisimpiä neuroverkkoihin liittyviä käsitteitä, kuten neuronit ja konvoluutioneuroverkot. GANeja hyödynnetään jo monissa eri toiminnoissa, kuten kuvien generoimisessa ja erilaisissa täydennystehtävissä. GAN on hyvin monipuolinen neuroverkko- rakenne, joten siitä löytyy uutta tietoa ja sen tutkimus etenee huimaa tahtia.

Tutkimuksen tavoitteena oli tarkastella erilaisia ongelmia mitä neuroverkot kohtaavat ja miksi nämä ongelmat syntyvät. Näistä suurin osa on odotetusti erilaisia datapohjaisia ongelmia, joissa datan heikko laatu vaikuttaa negatiivisesti neuroverkon oppimiseen ja siten aiheuttaa ongelmia. Näihin ongelmiin esitän ratkaisuksi GAN-pohjaisia ratkaisuja, sillä koen ne tehokkaiksi ratkaisuiksi dataan liittyvissä ongelmissa.

GAN on hyvin tehokas neuroverkkorakenne, mutta se ei ole ”hopealuoti” eli se ei pysty ratkaisemaan kaikkia ongelmia. Mutta se antaa ratkaisuja osaan ongelmista, mitä kasvojentunnistusohjelmat kohtaavat. GAN on parhaimmillaan erilaisissa tiedon täydennystehtävissä, joissa se on osoittautunut erimuodoissaan hyvinkin tehokkaaksi. Tutkimuksen perusteella on uskottavaa, että GAN tulee parantamaan kasvojentunnistusta tulevaisuudessa ja sen kehityksen seuraaminen tulee olemaan kiinnostavaa.

Lähteet

Albawi, Saad, Tareq Abed Mohammed ja Saad Al-Zawi. 2017. “Understanding of a convolutional neural network”. Teoksessa *2017 International Conference on Engineering and Technology (ICET),* 1–6. Ieee.

Arjovsky, Martin, ja Léon Bottou. 2017. “Towards principled methods for training generative adversarial networks”. *arXiv preprint arXiv:1701.04862.*

Arjovsky, Martin, Soumith Chintala ja Léon Bottou. 2017. “Wasserstein generative adversarial networks”. Teoksessa *International conference on machine learning,* 214–223. PMLR.

Bello, Martin G. 1992. “Enhanced training algorithms, and integrated training/architecture selection for multilayer perceptron networks.” *IEEE Transactions on Neural networks* 3 (6): 864–875.

Chavdarova, Tatjana, ja François Fleuret. 2018. “Sgan: An alternative training of generative adversarial networks”. Teoksessa *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 9407–9415.

Creswell, Antonia, ja Anil Anthony Bharath. 2018. “Inverting the generator of a generative adversarial network”. *IEEE transactions on neural networks and learning systems* 30 (7): 1967–1974.

De Oliveira, Rodrigo, Ramon Cristian Fernandes Araújo, Fabrício Barros, Adriano Segundo, Ronaldo Zampolo, Wellington Fonseca, Victor Dmitriev ja Fernando Brasil. 2017. “A System Based on Artificial Neural Networks for Automatic Classification of Hydro-generator Stator Windings Partial Discharges”. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications* 16 (syyskuu): 628–645. https://doi.org/10.1590/2179-10742017v16i3854.

Fukushima, Kunihiko, ja Sei Miyake. 1982. “Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition”. Teoksessa *Competition and cooperation in neural nets,* 267–285. Springer.

Goodfellow, Ian, Yoshua Bengio ja Aaron Courville. 2016. *Deep Learning.* http://www.deeplearningbook.org. MIT Press.

Goodfellow, Ian J, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville ja Yoshua Bengio. 2014. "Generative adversarial networks". *arXiv preprint arXiv:1406.2661.*

Hopfield, John J. 1982. "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the national academy of sciences* 79 (8): 2554–2558.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou ja Alexei A Efros. 2017. "Image-to-image translation with conditional adversarial networks". Teoksessa *Proceedings of the IEEE conference on computer vision and pattern recognition,* 1125–1134.

Jain, Anil K, Jianchang Mao ja K Moidin Mohiuddin. 1996. "Artificial neural networks: A tutorial". *Computer* 29 (3): 31–44.

Kortylewski, Adam, Andreas Schneider, Thomas Gerig, Bernhard Egger, Andreas Morel-Forster ja Thomas Vetter. 2018. "Training deep face recognition systems with synthetic data". *arXiv preprint arXiv:1802.05891.*

Lacey, Griffin, Graham W Taylor ja Shawki Areibi. 2016. "Deep learning on fpgas: Past, present, and future". *arXiv preprint arXiv:1602.04283.*

Lawrence, Steve, C Lee Giles ja Ah Chung Tsoi. 1997. "Lessons in neural network training: Overfitting may be harder than expected". Teoksessa *AAAI/IAAI,* 540–545. Citeseer.

Marriott, Richard T, Sami Romdhani ja Liming Chen. 2020. "A 3D GAN for Improved Large-pose Facial Recognition". *arXiv preprint arXiv:2012.10545.*

Martinez, Brais, ja Michel F Valstar. 2016. "Advances, challenges, and opportunities in automatic facial expression recognition". Teoksessa *Advances in face detection and facial image analysis,* 63–100. Springer.

Nwankpa, Chigozie, Winifred Ijomah, Anthony Gachagan ja Stephen Marshall. 2018. "Activation functions: Comparison of trends in practice and research for deep learning". *arXiv preprint arXiv:1811.03378.*

Rosenblatt, Frank. 1958. “The perceptron: a probabilistic model for information storage and organization in the brain.” *Psychological review* 65 (6): 386.

Wang, Kunfeng, Chao Gou, Yanjie Duan, Yilun Lin, Xinhu Zheng ja Fei-Yue Wang. 2017. “Generative adversarial networks: introduction and outlook”. *IEEE/CAA Journal of Automatica Sinica* 4 (4): 588–598.

Yang, Huiyuan, Zheng Zhang ja Lijun Yin. 2018. “Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks”. Teoksessa *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),* 294–301. IEEE.