Datan augmentoinnin menetelmien vertailu kuvadatalla

TURUN YLIOPISTO
Tietotekniikan laitos
Pro gradu -tutkielma
Tietojenkäsittelytieteet
Marraskuu 2021
Jaakko Honkanen

Ohjaaja:
Timo Knuutila

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO
Tietotekniikan laitos
JAAKKO HONKANEN: Datan augmentoinnin menetelmien vertailu kuvadatalla
Pro gradu -tutkielma, 50 sivua
Tietojenkäsittelytieteet
Marraskuu 2021

Datan augmentointi on keino parantaa koneoppimismallin kykyä suoriutua tehtävästä, jota varten se on opetettu. Data-augmentoinnin pääasiallinen tarkoitus on parantaa mallin yleistyvyyttä eli kykyä suoriutua sellaisen datan käsittelemisestä, jota ei ole opetusdatassa.

Tässä tutkielmassa vertaillaan eräitä datan augmentointimenetelmiä kuvamuotoiseen dataan sovellettuna. Tutkielmassa vertaillaan keskenään tyylinsiirto neuroverkolla -menetelmää sekä generatiivisiin kilpaileviin verkostoihin (GAN) perustuvia menetelmiä. Lisäksi tutkitaan niiden menetelmien käyttöä perinteisten, yksinkertaisempien datan augmentointimenetelmien rinnalla.

Data-augmentointimenetelmiä vertaillaan sen perusteella, kuinka paljon ne parantavat tarkkuutta kuvien luokittelutehtävissä. Tutkielmassa tutkitaan, miten hyvin generatiivisiin kilpaileviin verkostoihin perustuva datan augmentointi ja tyylinsiirtoon perustuva datan augmentointi suoriutuvat suhteessa perinteisiin augmentointimenetelmiin, ja miten ne muuttavat tilannetta verrattuna siihen, että ei käytettäisi augmentointia. Tutkielmassa käytetty data rajoittuu enimmäkseen kasvokuviin.

Kokeissa mitattujen luokittelun tarkkuustulosten perusteella todetaan, että generatiivisilla kilpailevilla verkostoilla sekä tyylinsiirrolla tehdyllä datan augmentoinnilla saavutetaan joissakin tapauksissa pieniä, mutta merkittäviä parannuksia. Todetaan myös, että tutkielmassa käytettyyn dataan generatiivisilla kilpailevilla verkostoilla tehty augmentointi sopii yleensä paremmin kuin tyylinsiirto ja perinteiset datan augmentoinnin menetelmät.

Asiasanat: datan augmentointi, GAN, tyylinsiirto neuroverkolla

Sisällys

1. Johdanto

Datan augmentointi eli data-augmentointi eli aineiston täydennys on keino parantaa koneoppimismallin kykyä suoriutua tehtävästä, jota varten se on opetettu. Data-augmentointia käytetään usein erityisesti, kun käsitellään kuvamuotoista dataa. Data-augmentoinnin pääasiallinen tarkoitus on parantaa mallin yleistyvyyttä eli kykyä suoriutua sellaisen datan käsittelemisestä, jota ei ole opetusdatassa. Data-augmentointia voidaan käyttää myös muihin tarkoituksiin. (Shorten & Khoshgoftaar, 2019)

Tässä tutkielmassa vertaillaan eräitä datan augmentointimenetelmiä kuvamuotoiseen dataan sovellettuna. Tutkielmassa vertaillaan tyylinsiirto neuroverkolla -menetelmää (*neural style transfer*) sekä generatiivisiin kilpaileviin verkostoihin (*generative adversarial network*) eli GANeihin perustuvia menetelmiä perinteisiin, yksinkertaisempiin data-augmentointimenetelmiin.

Tavanomaisimmin käytetyt eli perinteiset datan augmentointimenetelmät perustuvat siihen, että datasetin kuvien perusteella luodaan augmentoinnissa käytettäviä kuvia yksinkertaisella kuvan käsittelyllä. Kuvia voidaan esimerkiksi muuttaa peilikuvaksi, skaalata suuremmaksi tai kiertää keskipisteensä ympäri.

Data-augmentointimenetelmiä vertaillaan sen perusteella, kuinka paljon ne parantavat tarkkuutta kuvien luokittelutehtävissä.

Data-augmentointia käytetään pääsääntöisesti tilanteissa, joissa dataa on vähän tai se on liian yksipuolista.

Data-augmentointia käytetään usein ylinäytteistykseen *(oversampling)*. Niin voidaan tehdä tilanteissa, joissa datasetti on jaettu luokkiin, ja joissakin luokissa on paljon vähemmän kuvia kuin toisissa. Silloin datasettiä voidaan täydentää kuvilla, jotka on tuotettu jollakin data-augmentointimenetelmällä. (Shorten & Khoshgoftaar, 2019)

Datan augmentointia voi käyttää myös parantamaan mallin yleistettävyyttä tilanteissa, joissa luokkien välillä ei ole epätasapainoa. (Shorten & Khoshgoftaar, 2019)

Sen lisäksi että datan augmentointia voidaan käyttää parantamaan koneoppimismallin yleistyvyyttä, sitä voidaan käyttää myös muihin tarkoituksiin. Esimerkiksi generatiivisilla kilpailevilla verkostoilla toteutetulla data-augmentoinnilla voidaan korvata alkuperäiset kuvat tietosuojan vuoksi. (Liu ym., 2019; Ma ym., 2020)

Tämän tutkielman tutkimuskysymyksiä ovat: miten hyvin generatiivisiin kilpaileviin verkostoihin perustuva datan augmentointi ja tyylinsiirtoon perustuva datan augmentointi suoriutuvat suhteessa perinteisiin, yksinkertaisempiin augmentointimenetelmiin, ja miten ne muuttavat tilannetta verrattuna siihen, että ei käytettäisi augmentointia? Näihin kysymyksiin vastataan edellä mainittujen data-augmentoinnin käyttötarkoitusten näkökulmasta.

Tutkielmassa käytetty data rajoittuu enimmäkseen kasvokuviin, mutta data-augmentointia kokeillaan hieman myös satelliittikuvilla.

Luvussa 2 kerrotaan tarkemmin data-augmentoinnin käyttötarkoituksista. Luvussa 3 esitellään erilaisia datan augmentoinnin menetelmiä. Luvussa 4 kerrotaan, millä tavalla tässä tutkielmassa mitattiin augmentointimenetelmien vaikutusta luokitteluun. Luvussa 5 esitellään mitatut tulokset ja pohditaan niiden tulkintoja. Luku 6 koskee tutkielman rajoitteita ja tulosten yleistyvyyttä. Luvussa 7 kerrataan olennaisimmat tulokset ja käsitellään sitä, miten ne vertautuvat aiheesta tehtyihin tutkimuksiin.

2. Datan augmentoinnin käyttötarkoitukset

2.1. Datan augmentointi

Datan augmentoinnin tavoitteena on täydentää koneoppimismallin opetukseen käytettävää dataa generoimalla keinotekoista dataa oikean datan pohjalta (Shorten & Khoshgoftaar, 2019). Data-augmentoinnin tavoitteena on lisätä opetusdatan monimuotoisuutta (Zheng ym., 2019). Tällä tavalla voidaan vähentää ylisovittamista *(overfitting)* tilanteessa, jossa dataa on vähän, tai se on esimerkiksi painottunut niin että määrätyillä luokilla ei ole tarpeeksi opetusdataa luokitustehtävää varten. Datan augmentointi on siis tapa parantaa mallin yleistettävyyttä. (Shorten & Khoshgoftaar, 2019)

Datan augmentoinnilla voidaan mahdollisesti saada lisää tietoa irti alkuperäisestä datasetistä. (Shorten & Khoshgoftaar, 2019)

Datan augmentointia varten generoitujen kuvien on oltava sellaisia, että niissä on jäljellä ne alkuperäisten kuvien olennaiset asiat, joita koneoppimistehtävässä tarvitaan. Jos esimerkiksi tarkoituksena on luokitella kuvia, augmentointikuvan on kuuluttava samaan luokkaan kuin alkuperäinen kuva. Tämä vaatimus rajoittaa sitä minkälaisia augmentointimenetelmiä voidaan käyttää kussakin datasetissä. Esimerkiksi jos halutaan luokitella käsin kirjoitettuja numeroita, ei voida generoida lisää kuvia kääntämällä alkuperäisiä kuvia ylösalaisin, koska silloin numero saattaa muuttua toiseksi numeroksi, tai symboliksi, joka ei ole mikään numero. (Shorten & Khoshgoftaar, 2019)

Esimerkiksi lääketieteessä ei usein ole paljon dataa saatavilla koneoppimista varten, koska ihmisten yksityisyyden suoja rajoittaa tietojen saatavuutta. Muilla sovellusalueilla datan saatavuutta rajoittaa se, että datan merkitseminen *(labeling)* on työlästä ja aikaa vievää. Niissä tilanteissa datan augmentointia voidaan käyttää generoimaan lisää dataa. (Shorten & Khoshgoftaar, 2019)

On mahdollista, että oikeanlaisella datan augmentoinnilla voidaan parantaa koneoppimismallin suoriutumista silloinkin, kun dataa on saatavilla paljon (Wang & Perez, 2017).

2.2. Alkuperäisen datan korvaaminen ja täydentäminen koneoppimismallin tarkkuuden parantamiseksi

Datan augmentointi voidaan toteuttaa joko niin, että laajennetaan datasettiä muutetuilla tai generoiduilla kuvilla, tai niin, että käytetään muutettuja kuvia alkuperäisten kuvien sijaan. Näitä tekniikoita voidaan myös käyttää samaan aikaan, niin että koko alkuperäinen datasetti on korvattu muunnetuilla kuvilla, joita on enemmän kuin alkuperäisessä datasetissä. (Shorten & Khoshgoftaar, 2019)

Tässä tutkielmassa käsitellään molempia tekniikoita.

Data-augmentointi voidaan toteuttaa siten, että koneoppimismallin opetuksen aikana jokaisella erällä *(batch)* generoidaan uusia kuvia. (Krizhevsky ym., 2012) Data-augmentointi voidaan toteuttaa myös niin, että käytetään samoja kuvia koko opetuksen aikana (Wei ym., 2020).

Alkuperäisen datan täydentäminen augmentointidatalla on eräs ylinäytteistystekniikka. Sitä voidaan siis käyttää tilanteessa, jossa yhteen luokkaan kuuluvia datapisteitä on paljon vähemmän kuin toiseen luokkaan kuuluvia datapisteitä. Sellaisissa tilanteissa luokittelun tarkkuus saattaa kärsiä. Data-augmentoinnilla voidaan generoida lisää vähemmistöluokan kuvia niin että luokat eivät ole epätasapainossa. (Shorten & Khoshgoftaar, 2019)

2.3. Alkuperäisen datan korvaaminen tietosuojan vuoksi

Yksi tapa käyttää data-augmentointia on korvata sillä alkuperäinen data kokonaan siitä syystä, että alkuperäistä dataa ei haluta käyttää esimerkiksi tietosuojan vuoksi. Esimerkiksi lääketieteellinen tai yksilön taloudelliseen tilanteeseen liittyvä data voi olla arkaluontoista. Lisäksi henkilötietojen jakoa on yleisesti rajoitettu lainsäädännöllä. (Liu ym., 2019; Ma ym., 2020)

Eräs siihen tarkoitukseen soveltuva augmentointimenetelmä on generatiiviset kilpailevat verkostot. Generatiivisia kilpailevia verkostoja voidaan käyttää tuottamaan paljon korkealaatuisia kuvia siten, että ne eivät vaaranna ihmisten yksityisyyttä samalla tavalla kuin jos todellista dataa käytettäisiin sellaisenaan. Jos siis todellista opetusdataa ei voida esimerkiksi jakaa eteenpäin, on kuitenkin mahdollista, että esimerkiksi niillä opetettu generatiivinen kilpaileva verkosto voitaisiin jakaa. Toisaalta silloin olemassa riski, että

alkuperäisiä kuvia voidaan palauttaa verkoston painoista, jos sitä ei erikseen yritä estää verkoston suunnittelussa tai opetusprosessissa. (Liu ym., 2019; Ma ym., 2020)

3. Datan augmentoinnin menetelmät

3.1. Tyylinsiirto

3.1.1. Tyylinsiirto yleisesti

Tyylinsiirto on menetelmä, jossa erotetaan kuvista automaattisesti tyyli ja sisältö, ja yhdistetään ne uudeksi kuvaksi, jossa yhteen kuvaan on sovellettu toisen tyyliä. Tyylinsiirron avulla voidaan esimerkiksi saada valokuva näyttämään maalaukselta. (Gatys ym., 2015)

Nykyiset tyylinsiirtoalgoritmit voidaan jakaa kahteen ryhmään. Ensimmäisen ryhmän algoritmit toimivat siten, että tuotettava tyylin ja sisällön yhdistelmäkuva on aluksi pelkkää kohinaa, ja sitä muutetaan vähä vähältä sellaiseksi, että siinä on tyylikuvan tyyli ja sisältökuvan sisältö. Toisen ryhmän algoritmit toimivat siten, että opetetaan neuroverkko tuottamaan kuvia, joilla on määrätty tyyli. Opetuksen jälkeen neuroverkolle voidaan antaa syötteenä sisältökuva, jolloin neuroverkko tulostaa kuvan, jossa syötekuvaan on sovellettu tyyliä. Jälkimmäiset algoritmit ovat ainakin opetuksen jälkeen nopeampia. (Zheng ym., 2019)

Alkuperäinen tyylinsiirtoalgoritmi kuuluu ensimmäiseen ryhmään. Alkuperäinen tyylinsiirtoalgoritmi toimii näin:

Tyylinsiirtoon tarvitaan kaksi kuvaa: tyylikuva (*style image*) ja sisältökuva (*content image*). Tyylikuvasta erotetaan tyyli, ja sisältökuvasta sisältö, ja ne yhdistetään kohdekuvaksi (*target image*). Sitä varten määritellään kaksi virhefunktiota (*loss function* eli häviöfunktio, hukkafunktio tai tappiofunktio)**:** tyylivirhe ja sisältövirhe.

Sisältövirhe perustuu siihen, että otetaan konvoluutioneuroverkosta *(convolutional neural network)* syvemmistä verkon kerroksista *(layer)* painoja. Ne edustavat kuvasta löydettyjä korkean tason piirteitä. Korkean tason piirteissä on jätetty pois yksityiskohdat kuten tekstuuri, ja jätetty jäljelle esimerkiksi muotojen ääriviivat.

Sisältövirhe perustuu kohde- ja sisältökuvien erotukseen:

$$L_c = \frac{1}{2}\sum(T_c - C_c)^2$$

missä T_c ja C_c ovat kohde- ja sisältökuva.

Tyylivirhe perustuu sisältö- ja tyylikuvien Gram-matriisien *(Gram matrix)* eroon. Tyylivirhe lasketaan kaavalla

$$L_s = a\sum_i w(T_{s,i} - S_{s,i})^2$$

missä

- $T_{s,i}$ on kohdekuvan Gram-matriisi joka on laskettu kerroksessa i
- $S_{s,i}$ on tyylikuvan Gram-matriisi joka on laskettu kerroksessa i
- w_i on kerroskohtainen paino. Kerroskohtaiset painot ovat mallin hyperparametreja.
- a on kerroin, jota käytetään normalisoimiseen

Virhefunktiota varten lasketaan Gram-matriisi. Gram-matriisi G lasketaan kaavalla $g_{ij} = \mathbf{v}_i^T \mathbf{v}_j$. Sitä käytetään mittaamaan samankaltaisuutta eri kuvien välillä tyylin suhteen. Eli jos kahden kuvan Gram-matriisien erotusten itseisarvo on pieni, kuvia pidetään tyyliltään samankaltaisina. Gram-matriisin laskemista varten verkon määrätyn kerroksen piirrekartta *(feature map)* litistetään yksiulotteiseksi.

Kohdekuva tuotetaan siten, että minimoidaan sisältö- ja tyylivirheen painotettu summa.

Tyylinsiirtoa varten tarvitaan neuroverkko, joka on opetettu siten, että se tunnistaa sekä sisällön että tyylin erottamiseen tarvittavia piirteitä. Esimerkiksi voidaan käyttää kuvien luokitteluun opetettua neuroverkkoa, joka on opetettu suurella datasetillä. Neuroverkosta on valittava kerrokset, joiden tulosteesta *(output)* lasketaan sisältö- ja tyylivirhe.

3.1.2. Tyylinsiirto datan augmentoinnissa

Datan augmentointi tyylinsiirrolla voi olla hyödyllistä erityisesti tilanteissa, joihin monet perinteiset data-augmentoinnin menetelmät eivät sovellu. Esimerkiksi jos täytyy arvioida kuinka kaukana valokuvan eri kohdat ovat kamerasta, ei voida soveltaa menetelmiä, jotka muuttaisivat kuvan osien kokoa, koska esineiden koko on olennainen tieto etäisyyden arviointia varten. Ei myöskään ole hyötyä soveltaa satunnaisen alueen

poistamismenetelmää, koska ei ole tarvetta arvioida etäisyyttä sellaiseen kuvan kohtaan, jota ei näy kamerassa. (Jackson ym., 2018)

Tyylinsiirto voi olla hyödyllinen menetelmä sellaisten robottien konenäön parantamiseen, jotka toimivat vaihtelevassa ympäristössä, jonka kaikkia ominaisuuksia ei ole riittävästi edustettuna opetusdatassa, ja jossa virheiden seuraukset ovat kalliita (Shorten & Khoshgoftaar, 2019). Esimerkiksi itseajavien autojen konenäköä varten tarvitaan koulutusdataa poikkeuksellisen huonoista sääolosuhteista, mutta saatavilla oleva data painottuu tavanomaisiin ja suotuisiin sääolosuhteisiin. Muuntamalla hyvällä kelillä otettuja kuvia huonon kelin kuviksi tällaista dataa voitaisiin generoida. (Wang & Perez, 2017) Tyylinsiirtoa voidaan käyttää myös siten, että generoidaan kuvadataa 3d-mallinnetusta ympäristöstä, ja tehdään siitä todenmukaisempaa soveltamalla tyylinsiirtoa. (Shorten & Khoshgoftaar, 2019)

Datan augmentointi tyylinsiirrolla lisää todennäköisyyttä, että opetettava neuroverkko ei kiinnitä huomiota pelkästään tekstuureihin, vaan myös muotoihin. (Jackson ym., 2018)

Tyylinsiirtoa data-augmentoinnin menetelmänä on tutkittu eräissä tutkimuksissa (Jackson ym., 2018; Wang & Perez, 2017; Zheng ym., 2019).

Esimerkiksi Zhengin ym. (2019) tutkimuksessa havaittiin, että tyylinsiirto paransi luokittelutarkkuutta kahdella prosenttiyksiköllä VGG16-neuroverkolla Caltech 101- ja Caltech 256 -dataseteillä. Lisäksi tutkimuksessa havaittiin, että tyylinsiirto paransi luokittelutarkkuutta enemmän kuin tutkimuksessa käytetyt perinteiset augmentointimenetelmät, ja tyylinsiirto yhdistettynä perinteisiin augmentointimenetelmiin tuotti parempia luokittelutuloksia kuin pelkkä tyylinsiirto.

Wangin ja Perezin tutkimuksessa (2017) havaittiin, että tyylinsiirto paransivat kuvien luokittelutarkkuutta hieman verrattuna siihen, että augmentointimenetelmiä ei käytettäisi. Kuitenkin perinteiset augmentointimenetelmät ali geometriset muunnokset ja värisävyjen muunnokset paransivat luokittelutarkkuutta enemmän.

3.2. Generatiiviset kilpailevat verkostot

3.2.1. Generatiiviset kilpailevat verkostot yleisesti

Generatiivinen kilpaileva verkosto (*generative adversarial network* eli GAN) on neuroverkkoarkkitehtuuri, jolla voidaan opettaa malli, jolla voidaan generoida dataa.

Tarkoitus on, että malli oppii opetusdatan jakauman piirreavaruudessa, jotta malli voi sen jälkeen generoida saman kaltaista dataa. GANeja käytetään yleensä kuvamuotoisen datan generoimiseen.

GAN koostuu kahdesta neuroverkosta, joita kutsutaan generaattoriksi *(generator)* ja diskriminaattoriksi *(discriminator)*. Diskriminaattori opetetaan tunnistamaan, onko sille syötetty kuva todellinen opetuskuva vai generaattoriverkon generoima kuva. Generaattori opetetaan tuottamaan sellaisia kuvia, että diskriminaattori ei osaa tunnistaa, ovatko ne todellisia vai generaattorin generoimia kuvia.

Generaattorille annetaan syötteenä satunnainen vektori, joka otetaan esimerkiksi normaalijakaumasta, ja generaattori tuottaa sen perusteella kuvan.

Generatiivisista kilpailevista verkostoista on kehitetty useita eri laajennuksia. Yksi niistä on DCGAN eli Deep Convolutional GAN (Radford ym., 2015). DCGANin tarkoitus on lisätä generaattoriverkon kompleksisuutta, ja projisoida *(project)* kuva niin että siinä on enemmän ulottuvuuksia, ja sitten käyttää dekonvoluutiokerroksia *(deconvolutional layer)* siihen että tehdään moniulotteisesta tensorista tulostekuva. Dekonvoluutiokerrokset tekevät kuvasta suuremman.

DCGANin olennaisia ominaisuuksia ovat:

- DCGAN on sellainen GANin kaltainen verkosto, jossa käytetään kerroksia, jotka tekevät konvoluutio-operaatioita. Tavallisessa GANissa käytetään monikerroksisia perseptroniverkkoja *(multilayer perceptron)* (Shorten & Khoshgoftaar, 2019)
- Diskriminaattorin konvoluutiokerrokset tuottavat pienemmän kuvan kuin mitä niille annetaan syötteenä, koska konvoluutioikkuna hyppää joidenkin pikselien yli.
- Aktivaatiofunktiona käytetään LeakyReLU-funktiota.
- Neuroverkko normalisoi jokaisen erän *(minibatch)* syötteet muuttamalla niitä kyseisen erän syötteiden keskiarvon ja keskihajonnan perusteella.

DCGAN-arkkitehtuuria käytetään, koska se voi generoida paremmin verrattain suuria kuvia kuin perinteinen GAN. DCGANilla voi generoida esimerkiksi 64x64 pikselin

kokoisia värikuvia, kun taas tavallisella GANilla on vaikeaa saada tuotettua niin suuria laadukkaita kuvia.

3.2.2. Generatiiviset kilpailevat verkostot datan augmentoinnissa

Generatiivisilla kilpailevilla verkostoilla generoituja kuvia voidaan käyttää augmentoimaan dataa.

GAN-augmentoinnilla voi saada lisää tietoa irti datasetistä (Shorten & Khoshgoftaar, 2019; Bowles ym., 2018). Voidaan esimerkiksi muuttaa luoda vähemmistöluokan datapisteitä enemmistöluokan datan perusteella, jos datan epätasainen jakautuminen luokkiin on ongelma. Lisäämällä GANeilla luotuja kuvia opetusdataan voidaan myös parantaa yleistettävyyttä samasta syystä, kuin esimerkiksi värejä muuttamalla tai kuvaa rajaamalla voidaan pakottaa neuroverkko kiinnittämään huomiota asioihin, joihin se ei muuten kiinnittäisi huomiota (Shorten & Khoshgoftaar, 2019). Generatiivisten kilpailevien verkostojen avulla voi saada irti lisää tietoa datasetistä generoimalla kuvia, jotka muistuttavat oikeita kuvia. (Bowles ym., 2018)

GANien käyttöä datan augmentoinnissa rajoittaa se, että niillä voidaan tuottaa hyvälaatuisia kuvia vain, jos kuvat ovat pieniä. Sellaisia ovat esimerkiksi MNIST-datasetin kuvat, jotka ovat harmaasävyisiä 28x28 pikselin kokoisia kuvia, jotka esittävät käsin kirjoitettuja numeroita. Sen sijaan esimerkiksi ImageNet-datasetin kuvat ovat 250 pikseliä leveitä ja korkeita, ja niissä on kolme värikanavaa, ja ne ovat liian suuria, jotta GANilla voisi käytännössä generoida samanlaisia kuvia. (Shorten & Khoshgoftaar, 2019)

Isompien kuvien generointiin sopivat paremmin DCGAN- tai Progressively Growing GAN -arkkitehtuurit. (Shorten & Khoshgoftaar, 2019)

DCGAN voi luoda kuvia, jotka ovat kooltaan 64x64 kolmella värikanavalla. Siten se voi luoda suurempia kuvia kuin perinteinen GAN. (Shorten & Khoshgoftaar, 2019)

Toinen korkeamman resoluution kuvia tuottava GAN on Progressive Growing GAN. Progressive Growing GAN opettaa sarjan neuroverkkoja, jossa seuraava neuroverkko tuottaa suuremman resoluution kuvan kuin edellinen. Sen sijaan että neuroverkoille syötettäisiin satunnaislukuja sisältävä vektori, niille syötetään edellisen neuroverkon tuottaman kuva. Tällä menetelmällä on tuotettu erittäin korkealaatuisia kasvokuvia. (Shorten & Khoshgoftaar, 2019)

CycleGAN parantaa tuotettujen kuvien laatua, mutta ei välttämättä resoluutiota verrattuna tavalliseen GANiin. Siihen kuuluu cycle-consistency loss -funktio, joka auttaa stabilisoimaan GANia opetuksen aikana. Siinä missä tyylinsiirto oppii muunnoksen yhdestä kuvasta toiseen, CycleGAN oppii muunnoksen yhdestä sovellusalueesta (*domain*) toiseen sovellusalueeseen, kuten seeproista hevosiin. Generaattori ottaa hevosten kuvia ja tekemään niistä seeprojen kuvia siten, että diskriminaattori ei osaa sanoa kumpaan joukkoon ne kuuluivat alun perin. Tämän jälkeen generoidut seeprojen kuvat syötetään sellaisen verkon läpi, joka muuttaa ne takaisin hevosiksi. Toinen diskriminaattori määrittelee, kuuluuko tämä uudelleen muunnettu kuva hevosiin vai ei. Molempien diskriminaattorien virhearvot *(loss)* yhdistetään yhdeksi cycle-consistency -virheeksi. (Shorten & Khoshgoftaar, 2019)

CycleGAN sopii tilanteisiin, joissa yhden luokan kuvia voidaan käyttää toisen luokan kuvien generoimiseen (Wei ym., 2020). Esimerkiksi Wein ym. (2020) tutkimuksessa käytettiin CycleGANia siihen, että harmittomia polyyppeja esittävien kuvien pohjalta tehtiin syöpää enteilevien polyyppien kuvia. Sillä tavalla parannettiin luokittajan kykyä tunnistaa syöpää enteilevät polyyppejä. Tämä oli tarpeellista, koska syöpää enteilevien polyyppien kuvia on saatavilla paljon vähemmän kuin harmittomien polyyppien kuvia, ja siten opetuksessa käytettävät luokat ovat epätasapainoiset.

Wei ym. (2020) myös käyttivät luokittelijaa valitsemaan CycleGANilla generoiduista kuvista ne, jotka luultavasti olivat keskeisimpiä esimerkkejä siitä luokasta, jota ne edustivat. Luokittaja opetettiin kaikilla luokan esimerkeillä. Kun luokittajalle annetaan kuva, se tulostaa todennäköisyyden, että kuva kuuluu määrättyyn luokkaan. Valittiin vain ne luokkaan kuuluvat kuvat, jotka kuuluvat luokittajan mukaan siihen luokkaan suurella varmuudella. Vain niitä kuvia käytettiin syötteenä generatiiviselle kilpailevalle verkostolle. Siten parannettiin CycleGANin tuottamien kuvien laatua.

Ehdollinen GAN *(conditional GAN)* on myös eräs arkkitehtuuri, joka voi olla hyödyllinen datan augmentoimista varten (Shorten & Khoshgoftaar, 2019). Ehdollisissa GANeissa käytetään hyödyksi datasetissä olevia luokkia, mikäli datasetti on jaettu luokkiin. Tieto luokista syötetään neuroverkkojen ylimääriseen syötekerrokseen. Luokkien sijaan voidaan käyttää myös muuta tietoa, joka kertoo jotakin datapisteestä. Ehdollisen GANin etu on, että ehdollisen GANin generaattorineuroverkolla voidaan generoida haluttuun

luokkaan kuuluva kuva, kun sille syötetään satunnaislukuvektorin lisäksi tieto luokasta. (Mirza ja Osindero, 2015)

Vaihteleva autoenkoodaaja -neuroverkkoja voi myös käyttää siihen, että GANilla saa generoitua tiettyyn haluttuun luokkaan kuuluvia kuvia. (Shorten & Khoshgoftaar, 2019)

Vaihteleva autoenkoodaaja -neuroverkko *(variational auto-encoder)* oppii piirteiden todennäköisyysjakauman data-avaruudessa, jossa datapisteet sijaitsevat. Sen perusteella voidaan tuottaa kuvia, joissa on todennäköisyysjakaumasta otetut piirteet. Enkoodaaja siis muuttaa syöteavaruuden kuvat pieniulotteiseen muotoon, ja dekoodaaja *(decoder)* muuttaa pieniulotteisen esityksen takaisin syöteavaruuteen. Muuttamalla todennäköisyysjakaumaa voidaan generoida kuvia, joilla on halutut piirteet. GANia voidaan käyttää vaihtelevan autoenkoodaajan kanssa niin, että dekoodaaja on GANin generaattoriosa, ja generaattorin syöte annetaan diskriminaattorille, joka oppii autoenkoodaajan virhefunktion *(loss function)*. (Larsen ym., 2016) Larsenin ym. (2016) tutkimuksessa generoidaan CelebA-datasetin alkuperäisten kasvokuvien pohjalta samanlaisia kuvia, joissa kuvassa esiintyvälle henkilölle on esimerkiksi lisätty silmälasit tai jossa henkilö on muutettu kaljuksi. Tällä tavalla voidaan augmentoida datasettiä sellaisilla piirteiden yhdistelmillä, jotka esiintyvät liian harvoin alkuperäisessä datasetissä, ja parantaa siten esimerkiksi luokittelun tarkkuutta (Shorten & Khoshgoftaar, 2019). Vaihtelevia autoenkoodaajia voidaan käyttää data-augmentointiin myös siten, että tuotetaan lisää samaan luokkaan kuuluvia kuvia. Vaihtelevien autoenkoodaajien huonona puolena datan augmentoinnissa on se, että sillä tavalla tuotetut kuvat saattavat olla sumeita. (Madani yms., 2018)

GANit tarvitsevat paljon dataa opettamiseen, joten datasetistä riippuen eivät ole kelvollinen ratkaisu. (Shorten & Khoshgoftaar, 2019)

Kuten aiemmin mainittiin, GAN-kuvien käyttäminen datan augmentoinnissa voi parantaa yksityisyyttä, jos opetuksessa käytetään pelkästään GANilla generoituja kuvia. (Liu ym., 2019; Ma ym., 2020)

Yleisin tilanne, jossa GANia käytetään data-augmentoinnissa, on sellainen, jossa on vähän dataa tai data on epätasapainottunut eri luokkien kesken. (Tanaka & Aranha, 2019)

Bowlesin ym. (2018) tutkimuksessa tutkittiin GANiin perustuvan data-augmentoinnin käyttöä aivokuvien segmentoinnissa. Siinä käytettiin Progressive Growing GANeja, koska niiden avulla voi generoida suuria kuvia. Lisäksi Progressive Growing GAN ei ole niin herkkä hyperparametrien vaihtelulle kuin jotkut muut generatiiviset kilpailevat verkostot. Tutkimuksessa opetettiin GAN joukolla opetuskuvia. Sen jälkeen GANilla tuotetut kuvat yhdistettiin samojen opetuskuvien kanssa datasetiksi, jolla opetettiin segmentoijaneuroverkko. Tutkimuksen mukaan sekä GAN että perinteiset data-augmentointimenetelmät useimmiten paransivat tuloksia, ja GAN ja perinteiset menetelmät samaan aikaan käytettynä paransivat tuloksia verrattuna siihen, että käytettiin vain toista niistä.

Kyseisessä tutkimuksessa havaittiin lisäksi, että GANilla tuotettujen kuvien käyttö data-augmentoinnissa paransi tuloksia eniten silloin, kun alkuperäistä opetusdataa oli saatavilla vähän. Jos todellista opetusdataa oli kuitenkin todella vähän, tulokset olivat huonompia kuin niissä tapauksissa, joissa sitä oli enemmän. Jos opetusdataa oli paljon, GANilla generoitujen kuvien lisääminen kuvien segmentoijaneuroverkon opetusdataan heikensi joissakin tapauksissa tuloksia. (Bowles ym., 2018)

Madani ym. (2018) tutkivat rintakehän röntgenkuvista koostuvan datan augmentointia. Tutkimuksessa käytettiin 2000 kuvaa normaalista rintakehästä, ja 2000 sydän- ja verisuonitaudeista kärsivän henkilön kuvaa, ja tavoitteena oli parantaa luokittelua näihin kahteen luokkaan. Kumpaakin luokkaa varten opetettiin GAN, jolla pystyttiin tuottamaan sen luokan kuvia. Niitä lisättiin alkuperäiseen dataan noin neljäsosan verran alkuperäisen datan määrästä. Tutkimuksen mukaan GANit kykenivät tuottamaan realistisia kuvia. GANilla tuotetuilla kuvilla augmentoidulla datalla saatiin suurempi luokittelutarkkuus verrattuna siihen, että käytetään pelkkiä perinteisiä augmentointimenetelmiä, tai että ei käytetä data-augmentointia. Parannus luokittelutarkkuuteen oli tutkimuksen mukaan merkittävä, mutta ei suuri.

Perinteiset datan augmentointimenetelmät tuottavat uusia kuvia vaihtamalla väriä ja rajaamalla, mutta sillä tavalla saattaa syntyä epärealistisia kuvia. GANeilla sitä vastoin voidaan ymmärtää datan piilevä rakenne, ja sen avulla tuottaa uusia realistisia kuvia. (Madani ym., 2018)

On hidasta ja työlästä merkitä lääketieteellisiin kuviin tietoja, mitä ne kuvat sisältävät. Kuitenkin valvottu syväoppiminen vaatii paljon dataa. Siksi generatiivisten mallien käyttö on hyödyllistä tilanteissa, joissa on paljon dataa, jonka sisältöä ei ole merkitty koneoppimisohjelmia varten. (Madani ym. 2018)

Datan augmentointi on tärkeää erityisesti silloin kun opetetaan konvoluutioneuroverkkoja. Se johtuu siitä, että konvoluutioverkon oppimat piirteet ovat riippuvaisia siitä minkälaisessa asennossa kuvassa oleva esine tai asia on. Konvoluutioverkko voi oppia rotaatiosta riippumattomia piirteitä vain siinä tapauksessa, että opetusdatassa on tarpeeksi eri asennossa olevia kohteita, joita kuva esittää. (Bowles ym., 2018)

Jos opetusdataa on vähän, on mahdollista, että opetusdatan kuvissa on epäolennaisia ominaisuuksia, jotka sattumalta korreloivat sen kanssa mihin luokkaan kuvat kuuluvat, ja luokittelijaneuroverkko oppii virheellisesti kiinnittämään huomiota niihin. Jos opetussettiin lisätään tarpeeksi suuri määrä GANilla tuotettuja kuvia, kyseiset epäolennaiset ominaisuudet muuttuvat kohinaksi, jossa ei ole erotettavissa merkittävää korrelaatioita luokkien kanssa. (Bowles ym., 2018)

Lähteen Bowles ym. (2018) mukaan ihanteellinen GAN muuttaa datasetin jakauman diskreetistä jatkuvaksi jakaumaksi.

Lähteen Bowles ym. (2018) mukaan perinteisten data-augmentointimenetelmien etu GANeihin verrattuna on se, että perinteiset menetelmät voivat ekstrapoloida dataa niin että jakaumaan tulee mukaan sellaisia ääripäitä, joita ei ole mukana alkuperäisessä datassa. GANeilla taas voi vain interpoloida niin että tuotetaan sellaisia kuvia, jotka ovat piirteiltään olemassa olevien datapisteiden välissä. Perinteiset augmentointimenetelmät ja GAN augmentointimenetelmänä voivat siis tuottaa erilaisia kuvia, ja siten täydentää toisiaan.

Lähteen Bowles ym. (2018) mukaan GANien ongelma data-augmentoinnin välineinä on, että usein ei voida generoida kuvia, jotka ovat yhtä laadukkaita kuin alkuperäiset. Toisaalta niiden ei tarvitse olla yhtä laadukkaita parantaakseen tuloksia. Augmentointidatan lisääminen usein parantaa tuloksia, mutta toisaalta jos data on huonolaatuista, se heikentää tuloksia. Kun harkitaan GANin käyttöä data-

augmentointimenetelmänä, olennainen kysymys on, kumpi näistä vaikutuksista on suurempi.

Generatiivisia kilpailevia verkostoja voi siis käyttää data-augmentointimenetelmänä siten, että tehdään yhden tyyppisistä kuvista toisen tyyppisiä kuvia esimerkiksi CycleGANilla (Sandfort ym., 2019; Madani ym., 2018). Sen lisäksi niitä voi käyttää tekemään uusia kuvia pelkästään saman tyyppisten kuvien pohjalta (Liu ym., 2019; Ma ym., 2020; Madani ym., 2018)

3.3. Muut augmentointimenetelmät

Menetelmät, joilla tuotetaan keinotekoisia kuvia olemassa olevien pohjalta, voidaan luokitella yhteen seuraavasta kahdesta kategoriasta: tavallinen kuvamanipulaatio ja syväoppimiseen perustuva kuvien generointi tai muuntaminen (Shorten & Khoshgoftaar, 2019). Tässä työssä käsiteltävistä menetelmistä tyylinsiirto ja generatiiviset kilpailevat verkostot edustavat jälkimmäistä kategoriaa ja perinteiset augmentointimenetelmät edellistä kategoriaa.

Perinteisiä augmentointimenetelmiä ovat värimuunnokset kuten värisävyn ja kirkkauden muunnokset, geometriset muunnokset kuten kuvan kääntäminen eli rotaatio tai peilikuvaksi muuttaminen, kuvan osan poistaminen tai korvaaminen toisella kuvalla, kohinan lisääminen kuvaan tai ydinsuotimien *(kernel filter)* käyttäminen. Perinteisistä menetelmistä kerrotaan tarkemmin luvussa 4.

Muita olemassa olevia datan augmentointimenetelmiä ovat esimerkiksi vihamielistä koneoppimista hyödyntävä opetus *(adversarial training)* ja metaoppiminen *(meta learning)*. Vihamielistä koneoppimista hyödyntävässä opetuksessa neuroverkko oppii sellaisia kuvien augmentointeja, jotka johtavat siihen, että luokittelijaverkko luokittelee kuvan väärin. Siten voidaan löytää luokittelijaverkon heikkoja kohtia. Metaoppimisessa menetelmänä on löytää automaattisesti käsiteltävään dataan sopivia data-augmentointimenetelmien yhdistelmiä. (Shorten & Khoshgoftaar, 2019) Vihamielistä koneoppimista hyödyntävää opetusta ja metaoppimista ei käsitellä tässä työssä.

Datan augmentoinnin soveltaminen validointidatasetin laajentamiseen *(test-time augmentation)* kuuluu myös datan augmentoinnin piiriin (Shorten & Khoshgoftaar,

2019), mutta tässä tutkielmassa käsitellään vain opetusdataan kohdistuvaa data-augmentointia.

4. Vertailussa käytettävät kriteerit ja koejärjestelyt

4.1. Vertailu luokittelutehtävissä

Data-augmentoinnin vaikutuksen mittaaminen tehtiin siten, että luokiteltiin kuvat koneoppimismallilla niin, että opetusdatassa käytettiin eri data-augmentointimenetelmillä generoitua dataa alkuperäisen datan lisäksi. Kirjoitettiin ylös luokittelun tarkkuutta kuvaavat arvot.

Kaikissa luokittelutehtävissä data jaettiin kahteen luokkaan sen perusteella, kuuluuko datapiste määrättyyn luokkaan vai ei. Luokittelu tehtiin muutamalle eri luokalle molemmassa datasetissä.

Käytetty luokittelija-neuroverkko on rakenteeltaan saman lainen kuin luvussa 4.3.2 kuvailtu diskriminaattoriverkko. Siinä käytetään kuitenkin optimointialgoritmina stokastinen laskeutuva gradientti -algoritmia *(stochastic gradient descent)*, jonka learning rate -parametrina on 0,001 ja momenttitermiparametrina *(momentum)* on 0,9. Virhefunktiona *(loss function)* käytetään binääristä ristientropiafunktiota *(binary cross-entropy)*.

Luokittelun tarkkuutta mitattiin seuraavilla mittareilla:

- täsmällisyys *(accuracy)* eli oikein luokiteltujen määrä / datapisteiden kokonaismäärä
- tarkkuus (*precision*) eli oikeat positiiviset / (oikeat positiiviset + väärät positiiviset)
- herkkyys *(recall* eli saanti*)* eli oikeat positiiviset / (oikeat positiiviset + väärät negatiiviset)
- F1-mitta *(f-score)* eli (2 * tarkkuus * herkkyys) / (tarkkuus + herkkyys)

4.2. Tutkielmassa käytetyt datasetit

Luokittelussa käytettiin CelebA-datasettiä (Liu ym., 2015) sekä EuroSAT-datasettiä (Helber ym. 2017). Nämä datasetit valittiin, koska ne molemmat ovat sellaisia, joihin voidaan soveltaa monia eri data-augmentointimenetelmiä. Jos sen sijaan datana käytettäisiin esimerkiksi käsin kirjoitettuja merkkejä, niiden tunnistuksessa muun muassa tyylinsiirto ei olisi sopiva tai hyödyllinen augmentointimenetelmiä. Kyseiset datasetit

valittiin myös sen vuoksi, että ne ovat suuria, ja tässä tutkielmassa käytettävät generatiivinen kilpaileva verkosto vaatii paljon opetusdataa. Lisäksi molemmat datasetit ovat jaettavissa helposti eri luokkiin ja ne ovat helposti saatavilla.

CelebA-datasetti koostuu julkisuuden henkilöiden kasvokuvista. Se on jaettu opetus-, testi- ja validointiosaan siten, että opetusdatasetissä on noin 160 000 kuvaa. Testi- ja validointisetissä on molemmissa noin 20 000 kuvaa. CelebA-datasettiin kuuluu tieto siitä mitä ominaisuuksia kuvissa on. Ominaisuuksia on yhteensä 40, ja jokaista kuvaa kohti on merkitty, onko ominaisuus läsnä siinä kuvassa vai ei. Ominaisuuksia ovat esimerkiksi silmälasit, musta hiusväri tai miessukupuoli. CelebA-datasetin kuvien koko on 218x178 pikseliä, ja ne ovat värikuvia. (Liu ym., 2015)

CelebA-datasetistä käytettävät luokat on lueteltu taulukossa 1.

Taulukko 1 Tässä tutkielmassa käytetyt CelebA-datasetin luokat.

Tunniste	Nimi	Kuvien lukumäärä opetusdatasetissä	Kuvien osuus opetusdatasetistä
Attribuutti 21	*Mouth slightly open* (suu hieman avoinna)	n. 78 000	n. 48 %
Attribuutti 8	*Black hair* (mustat hiukset)	n. 39 000	n. 24 %
Attribuutti 16	*Goatee* (pukinparta)	n. 10 000	n. 6 %
Attribuutti 35	*Wearing hat* (hattu päässä)	n. 8 000	n. 5 %

Data-augmentointimenetelmien vaikutusta selvitettiin myös EuroSAT-datasetillä. Sillä tehtiin suppeampi määrä kokeita.

Datan augmentointia kasvokuvilla on tutkittu aikaisemmin tutkimuksissa (Wang ym., 2019) ja (Lee ym., 2020). Leen ym. tutkimuksessa luotiin vähemmistöluokan kuvia enemmistöluokan kuvia käyttäen.

EuroSAT-datasetissä on 27 000 kuvaa. Ne on jaettu 10 luokkaan. Jokaisessa luokassa on 2 000–3 000 kuvaa. Kuvat ovat satelliittikuvia laajasta alueesta. Datasetti on jaettu luokkiin sen perusteella, onko kuvassa esimerkiksi metsää, peltoa vai teollisuusaluetta.

EuroSAT-datasetti jaettiin niin että 4/5 kustakin luokasta käytettiin opetussettinä ja 1/5 käytettiin testisettinä.

Kuvassa 1 on esimerkkejä EuroSAT-datasetin kuvista.

Kuva 1. Esimerkkejä EuroSAT-datasetin kuvista (Helber ym. 2017).

4.3. Tutkielmassa käytettyjen augmentointimenetelmien yksityiskohdat

4.3.1. Tyylinsiirron toteutus

Tässä tutkielmassa käytetään samaa tyylinsiirtomenetelmää, joka on kuvailtu tutkimuksessa (Jackson ym., 2018). Tässä tutkielmassa käytetään myös sen tutkimuksen yhteydessä julkaistua ohjelmakoodia ja tyyliä kuvaavia vektoreita, jotka on laskettu valmiiksi sitä tutkimusta varten. Kyseistä menetelmää käytetään, koska alkuperäinen tyylinsiirtoalgoritmi on liian hidas, ja koska kyseinen menetelmä ei vaadi erillistä datasettiä tyylikuvia varten, koska tyylivektorit on jo laskettu.

Lähde (Jackson ym., 2018) esittelee menetelmän, jolla voi soveltaa kuvaan satunnaista tyyliä. Menetelmä toimii siten, että on laskettu tyylivektorien todennäköisyysjakauma Painter By Numbers -datasetissä, ja siitä jakaumasta otetaan satunnainen 100-ulotteinen tyylivektori jokaista kuvaa varten, johon halutaan soveltaa tyylinsiirtoa. Sillä tavalla simuloidaan sitä, että otettaisiin Painter By Numbers -datasetistä satunnainen kuva, ja erotettaisiin siitä tyylivektori. Tämä tekniikka kuitenkin vaatii vähemmän laskentatehoa. Menetelmä sopii moniin tilanteisiin sovellusalueesta *(domain)* riippumatta. (Jackson ym., 2018)

Tämän tyylinsiirtomenetelmän etu datan augmentoinnissa on myös se, että koska tyylivektorien todennäköisyysjakauma on laskettu suuresta Paint By Numbers -

datasetistä, opetettavan neuroverkon on opittava suoriutumaan hyvin yleisesti kuvien tyylistä riippumatta, koska se ei voi ylisovittua *(overfit)* tiettyihin tyyleihin. (Jackson ym., 2018) Jos käytettäisiin vain muutamia tyylejä, tulos olisi luultavasti huonompi, vaikka tyylit olisivat keskenään erilaisia (Zheng ym., 2019).

Kuvassa 2 on esitetty esimerkkejä tyylinsiirrolla generoiduista kuvista.

Kuva 2. Esimerkkejä tyylinsiirrolla tuotetuista CelebA-datasetin (Liu ym., 2015) kuvista verrattuna alkuperäisiin kuviin.

EuroSAT-datasettiin tyylinsiirtoa on aiemmin sovellettu tutkimuksessa (Helber ym., 2019). Siinä tunnistettiin automaattisesti erilaiset alueet kuvasta, kuten maa ja vesi. Niihin sovellettiin erikseen eri tyylejä, ja sen jälkeen ne yhdistettiin takaisin yhdeksi kuvaksi. Tässä tutkielmassa sitä vastoin sovelletaan samaa tyyliä koko satelliittikuvaan.

4.3.2. Generatiivinen kilpaileva verkosto

Generatiivisella kilpailevalla verkostolla tuotetut kuvat tuotettiin DCGAN-verkostolla.

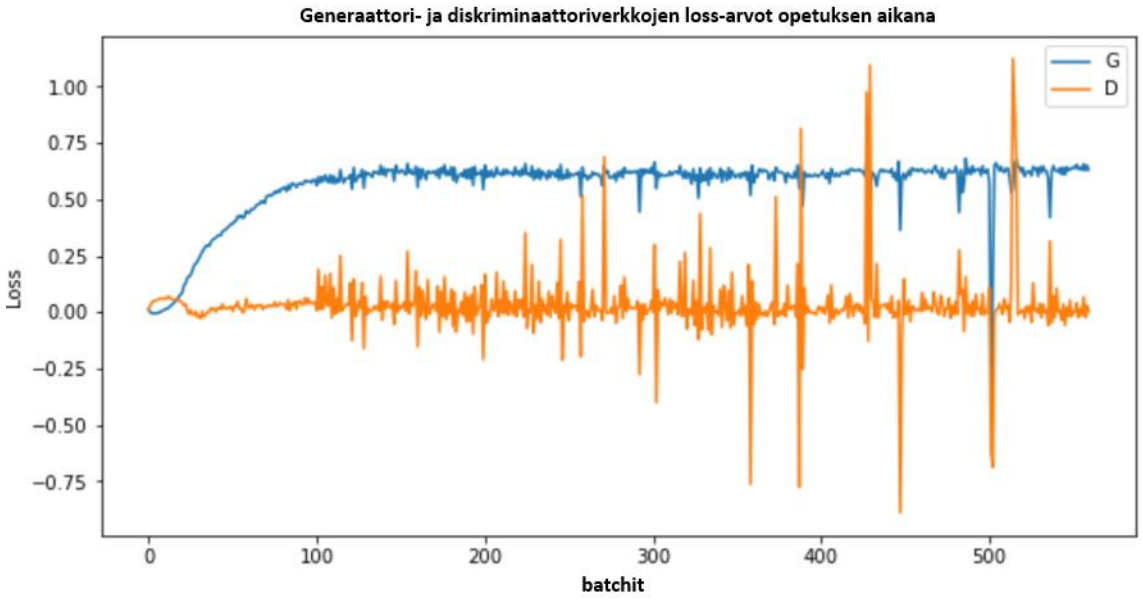
CelebA-datasetin kuvat rajattiin ja pienennettiin niin, että niiden korkeudeksi ja leveydeksi tuli 64. Sen jälkeen niiden pohjalta generoitiin uusia kuvia DCGAN-verkostolla.

Optimointialgoritmina käytettiin RMSprop-algoritmia, koska sitä suositellaan käytettäväksi WGAN-verkossa. Optimointialgoritmin learning rate -parametrina käytettiin lukua 0,00005. Samaa optimointialgoritmia ja parametria käytettiin sekä diskriminaattori- että generaattoriverkossa. Verkosto toteutettiin lähteessä DCGAN Tutorial (2021) esitetyn ohjelmakoodin pohjalta.

DCGAN-verkoston opettamista tehtiin viiden epookin *(epoch)* ajan. Jokaista epookkia kohti käytiin läpi 160 000 kuvan opetusdatasetti yhden kerran. Virhearvot *(loss)* eivät

muuttuneet johdonmukaisesti mihinkään suuntaan viimeisten epookkien aikana, joten opetuksen jatkamisella ei luultavasti olisi ollut vaikutusta.

Kuvassa 3 on virhefunktioiden tulosteiden kehitys opetuksen aikana, kun opetusdatan koko oli noin 40 000 kuvaa.



Kuva 3. Virhefunktion arvojen kehitys generaattori- ja diskriminaattoriverkkojen opetuksen aikana.

Käytetyssä diskriminaattorineuroverkossa on viisi konvoluutiokerrosta. Aktivointifunktiona jokaisessa kerroksessa käytetään LeakyReLU-funktiota. Toisen, kolmannen ja neljännen konvoluutiokerroksen kohdalla sovelletaan eränormalisointia *(batch normalization)*. Eränormalisointi tarkoittaa sitä, että normalisoidaan kerrokselle tulevat syötteet jokaisella erällä *(mini-batch)*. Syöte normalisoidaan siten että sen keskiarvo on 0 ja keskihajonta on 1 (Ioffe ja Szegedy, 2015).

Generaattoriverkossa taas on viisi kerrosta, joissa käytetään transponoitua konvoluutiota. Siinä sovelletaan myös eränormalisointia joidenkin konvoluutiokerrosten kohdalla. Aktivointifunktiona käytetään ReLU-funktiota.

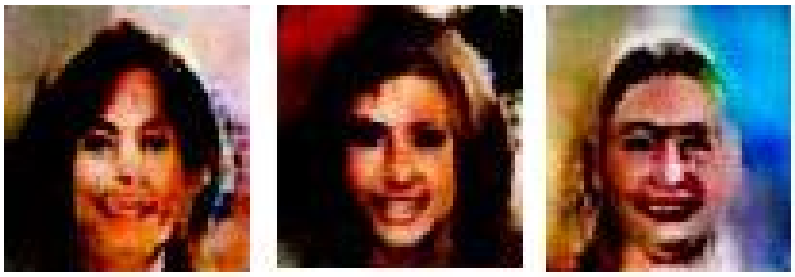
Diskriminaattori- ja generaattoriverkoissa kuvat esitetään niin että niillä on leveys, korkeus ja tietty määrä kanavia *(channel)*. Nämä ulottuvuudet ovat erilaiset eri kerroksissa. Diskriminaattorin syötekerroksessa on kolme kanavaa, koska värikuvissa on kolme värikanavaa. Syötekerroksen jälkeen tulevissa kerroksissa leveys ja korkeus ovat

pienempiä, kun taas kanavien määrä lisääntyy. Generaattoriverkossa taas korkeus ja leveys kasvavat kerros kerrokselta, ja kanavien määrä pienenee.

Perinteisessä GANissa diskriminaattori tuottaa luvun, joka on todennäköisyys, että diskriminaattorille syötetty kuva on todellinen opetuskuva. Tässä tutkielmassa diskriminaattori ei tuota todennäköisyyttä, vaan se antaa pistearvon, joka merkitsee kuvan todenmukaisuutta. Tavoitteena on opettaa diskriminaattori siten, että se tuottaa erilaisia pisteitä todellisille opetuskuville kuin generoiduille kuville. Diskriminaattorissa ja generaattorissa käytettiin Wasserstein loss -funktiota. Siten tässä tutkielmassa käytetty GAN on Wasserstein GAN eli WGAN (Arjovsky ym. 2017). WGANin diskriminaattoria sanotaan kriitikoksi (*critic*).

Tutkielmassa käytetään Wasserstein GANia, koska sillä tavalla voidaan välttää ilmiö, jossa verkko oppii generoimaan vain sellaisia kuvia, jotka ovat keskenään hyvin samankaltaisia (*mode collapse*).

Kuvassa 4 on esimerkkejä generatiivisilla kilpailevilla verkostoilla generoiduista kuvista.



Kuva 4. Esimerkkejä CelebA-datasetillä opetetulla GANilla generoiduista kuvista.

Lisäksi tehtiin joitakin kokeita CycleGANilla. Niissä opetusta tehtiin 200 epookin ajan. Tässä tutkielmassa käytetty CycleGAN perustuu toteutukseen lähteessä PyTorch-CycleGAN (2017), ja sen rakenne ja hyperparametrit ovat samanlaiset.

4.3.3. Perinteiset data-augmentointimenetelmät

Tässä aliluvussa luetellaan perinteisiä data-augmentointimenetelmiä ja perusteellaan, miksi niitä käytetään tai ei käytetä tässä tutkielmassa.

4.3.3.1 Satunnaisen alueen poistaminen

Satunnaisen alueen poistaminen -menetelmässä *(random erasing)* poistetaan satunnaisen kokoinen osa kuvasta, eli korvataan poistetut pikselit vakioarvoilla tai kohinalla. Usein

satunnaisen alueen poistamista sovelletaan vain osaan datasetin kuvista. Usein poistettava alue on nelikulmio. (Shorten & Khoshgoftaar, 2019)

On olemassa myös menetelmiä, joissa kuvan sisältö vaikuttaa siihen mikä alue poistetaan (Shorten & Khoshgoftaar, 2019).

Satunnaisen alueen poistamisen avulla opetettu malli yleistyy paremmin tilanteisiin, joissa kuvassa olevasta olennaisesta esineestä tai asiasta osa on peitetty. (Zhong ym., 2017) Satunnaisen alueen poistaminen myös auttaa varmistamaan, että esineen tunnistus ei perustu ainoastaan yhteen yksityiskohtaan, koska se yksityiskohta voidaan poistaa joistakin kuvista. (Shorten & Khoshgoftaar, 2019)

Satunnaisen alueen poistamisen hyperparametrit ovat seuraavat:

- poistettavan nelikulmion kohtisuorien sivujen pituuksien suhde toisiinsa *(aspect ratio*)
- poistettavan nelikulmion pinta-ala suhteessa koko kuvan pinta-alaan
- se osuus datasetin kuvista, johon satunnaisen alueen poistamista sovelletaan, jos sitä ei sovelleta kaikkiin kuviin
- poistetun alueen pikselien arvot

Muun muassa tutkimuksessa (Zhong ym., 2017) esitettyjen arvojen perusteella päätettiin seuraavat hyperparametrien arvot:

- Poistettavan alueen sivujen suhteeksi valitaan jokaista kuvaa kohti satunnainen luku lukujen 0,3 ja 3,3 väliltä.
- Poistettavan nelikulmion pinta-alan suhteeksi koko kuvan pinta-alaan valitaan jokaista kuvaa kohti satunnainen luku väliltä 0,02 ja 0,4.
- Satunnaisen alueen poistamista sovelletaan satunnaisesti valittuun puolikkaaseen datasettiä.
- Poistettu alue korvataan mustilla pikseleillä.

On mahdollista, että satunnaisen alueen poistaminen ei sovellu moniin tässä tutkielmassa valittuihin CelebA-datasetin luokkiin, koska jotkut luokat määräytyvät jonkin yksityiskohdan perusteella. Tällainen luokkajako on esimerkiksi se, onko kuvassa olevan henkilön suu auki vai kiinni. Siten olennaisen yksityiskohdan peittäminen tekee

luokittelusta vaikeaa. Jos taas joku epäolennainen osa kuvasta peitetään, siitä ei luultavasti ole hyötyä, koska luokittelija ei joka tapauksessa voisi käyttää sitä kuvan osaa hyödyksi.

Lisäksi satunnaisen alueen poistamisesta tekee CelebA-datasettiin vähemmän sopivan se, että satunnaisen alueen poistamisen tarkoitus on parantaa luokittelijan yleistyvyyttä kuviin, joissa olennainen kohta on osittain peitetty. CelebA-datasetin testidatassa on kuitenkin harvoin mitään kasvojen edessä, tai sitten sama kohta on peitetty monista kasvoista. Näin on esimerkiksi siinä tapauksessa, että silmät ovat aurinkolasien takana tai hiukset ja otsa ovat hatun alla.

4.3.3.2 Kohinan lisääminen

Kohinan lisääminen kuviin on eräs data-augmentointimenetelmä, joka voi parantaa merkittävästi kuvien luokittelutuloksia (Moreno-Barea ym., 2018). Kohinan lisääminen voi auttaa neuroverkkoa oppimaan piirteitä, jotka ovat vähemmän herkkiä kuvien vaihtelulle (Shorten & Khoshgoftaar, 2019).

Usein kohinan lisääminen toteutetaan siten että lisätään kuvan pikseleihin satunnaisia arvoja, jotka on otettu normaalijakaumasta (Shorten & Khoshgoftaar, 2019).

Tässä tutkielmassa satunnaiset arvot otetaan tasajakaumasta.

4.3.3.3 Värimuunnokset

Värimuunnosten tarkoituksena on yleensä vähentää valaistuksen ja värien vaihtelun vaikutusta luokittelijaan (Taylor & Nitschke, 2017). Siinä muutetaan kuvan värejä satunnaisella tavalla (Shorten & Khoshgoftaar, 2019).

Luokittelutarkkuuden parantamisen lisäksi värimuunnoksia voidaan käyttää pienentämään datasetin kokoa muuttamalla kuvat harmaasävykuviksi, jos värit eivät ole olennaisia koneoppimistehtävän kannalta (Shorten & Khoshgoftaar, 2019).

Värimuunnosten huono puoli augmentointimenetelmänä on, että joissakin tapauksissa värit ovat olennainen tieto kuvassa (Shorten & Khoshgoftaar, 2019).

Värimuunnokset ovat olennaisia tämän tutkielman kannalta, koska värimuunnokset augmentointimenetelmänä vaikuttaa saman kaltaiselta kuin tyylinsiirto, ja siten värimuunnokset voivat olla vaihtoehto tyylinsiirrolle tai toisin päin.

Värimuunnokset luultavasti sopivat kohtalaisen hyvin CelebA-datasetin kasvokuviin, koska muodot ovat luultavasti värisävyjä olennaisempia piirteiden tunnistuksessa, ja kuvien valaistus vaihtelee.

Tässä tutkielmassa käytetään värimuunnoksia, jossa värikylläisyyttä *(saturation)* ja värisävyä *(hue)* muutetaan satunnaisesti jokaisen kuvan kohdalla. Ne kaksi ovat siis hyperparametrit, jotka täytyy valita. Ne valitaan käyttämällä näiden kahden arvon eri yhdistelmiä ja mittaamalla luokittelun tarkkuus F1-mitan perusteella.

4.3.3.4 Kirkkauden muuttaminen

Kirkkauden muuttaminen on eräs data-augmentointimenetelmä, joka kuuluu värimuunnoksiin. Tässä tutkielmassa kirkkautta käsitellään erillisenä menetelmänä.

Kirkkauden muutoksilla on myös silmämääräisesti arvioituna samanlainen vaikutus kuviin joissain tapauksissa kuin tyylinsiirrolla, ja siksi se otetaan mukaan vertailuun.

Kirkkauden muutokset voidaan valita sen mukaan, mikä on yleisen valaistuksen tai kameran asetusten aiheuttama vaihtelu kirkkaudessa, tai luonnollinen vaihtelu niiden asioiden väreissä, joita kuva esittää. (Perez ym., 2018)

Tässä tutkielmassa kirkkauden muutoksen suuruus valitaan kokeilemalla eri arvoja ja mittaamalla luokittelun tulos niillä arvoilla augmentoidulla datalla.

4.3.3.5 Translaatio ja rajaus

Kuvan siirtäminen leveys- tai pystysuunnassa eli translaatio on eräs data-augmentointimenetelmä. Kuvan rajaus on saman kaltainen augmentointimenetelmä. (Shorten & Khoshgoftaar, 2019) Molemmissa kuviin jää mahdollisesti tyhjää tilaa, jos kaikkien kuvien on oltava vakiomuotoisia. Tyhjän tilan pikselit voidaan asettaa määrätyksi vakioarvoksi.

Kuvan siirtämisen tarkoituksena on varmistaa, ettei opetettava koneoppimismalli ole liian riippuvainen siitä, että tietyt piirteet sijaitsevat usein tietyssä paikassa kuvaa. (Shorten & Khoshgoftaar, 2019)

Tässä tutkielmassa ei käytetä kuvan siirtämistä tai rajausta augmentointimenetelmänä, koska ne eivät oletettavasti sovellu CelebA-datasetin kuviin. CelebA-datasetissä opetettava neuroverkko saattaa kiinnittää paljon huomiota siihen, missä kohtaa kuvaa

piirteet sijaitsevat, koska kaikki kasvot ovat suunnilleen samankokoisia ja ne ovat kuvan keskellä. Kuitenkin testidatassa on nämä samat ominaisuudet, joten mallin yleistyminen ei luultavasti paranisi testidatalla mitattuna, jos käytettäisiin menetelmiä, jotka vähentävät riippuvuutta piirteiden sijainnista.

4.3.3.6 Kuvan kääntäminen keskipisteensä ympäri

Kuvan kääntäminen keskipisteensä ympäri *(rotation)* on eräs augmentointimenetelmä, jota voidaan käyttää samoihin tarkoituksiin kuten translaatio ja rajaus, eli torjumaan sijaintiin liittyviä vääristymiä. (Shorten & Khoshgoftaar, 2019)

Kuvan kääntäminen keskipisteensä ympäri ei luultavasti sovellu CelebA-datasetin kuvien augmentointimenetelmäksi, koska kaikki kuvat ovat pystysuorassa.

4.3.3.7 Kuvien sekoittaminen keskenään

Kuvia sekoittamalla luodaan uusi kuva kahden tai useamman alkuperäisen kuvan pohjalta.

Uusi kuva voidaan muodostaa siten, että siihen otetaan kahden kuvan pikselien keskiarvot. Muita mahdollisia tapoja sekoittaa kuvia keskenään on esimerkiksi ottaa toisesta kuvasta vasen ja toisesta oikea puoli ja yhdistää ne. Kuvien sekoittamista tutkineissa tutkimuksissa on havaittu, että ne parantavat mallin yleistettävyyttä, vaikka ei ole aina intuitiivisesti selvää, minkä takia, toisin kuin monissa muissa data-augmentointimenetelmissä. (Shorten & Khoshgoftaar, 2019)

Kuvien sekoittaminen saattaa soveltua kasvokuviin, jos se soveltuu muuhun kuvadataan. Sitä ei kuitenkaan oteta mukaan tässä tutkielmassa tehtävään vertailuun, koska ei ole selvää miksi se parantaisi tulosta tai millä tavalla kuvien sekoittamisen hyperparametrit pitäisi valita.

4.3.3.8 Kuvan muuttaminen peilikuvakseen

Kuvan muuttaminen peilikuvaksi on eräs perinteinen data-augmentointimenetelmä. Monissa dataseteissä ainoastaan peilaaminen y-akselin suhteen on kelvollinen augmentointimenetelmä, koska ylösalaisin kuvassa olevia asioita ei tule todellisessa datassa vastaan eikä niitä tarvitse pystyä luokittelemaan. (Shorten & Khoshgoftaar, 2019)

Kaikki kasvokuvat CelebA-datasetissä ovat oikein päin pystysuunnassa, joten peilaus x-akselin suhteen ei ole sopiva augmentointimenetelmä. Useimmat kasvokuvat ovat varsin

symmetrisiä y-akselin suhteen, joten y-akselin suhteen peilaamisesta ei luultavasti ole paljon hyötyä. Siitä saattaisi kuitenkin olla hyötyä niiden kuvien luokittelussa, joissa kuvaa ei ole otettu suoraan henkilön edestäpäin. Peilaamista ei kuitenkaan käytetä tässä tutkielmassa.

4.3.3.9 Ydinsuotimet

Ydinsuotimia *(kernel filter)* käytetään kuvankäsittelyssä sumentamaan tai terävöittämään kuvia. Ydinsuotimet toimivat niin että konvoluutio-operaatiossa liu'utetaan kuvan yli matriisia, joka on joko sumentava tai terävöittävä suodin *(filter)*. (Shorten & Khoshgoftaar, 2019)

Kuvien sumentaminen data-augmentoinnin menetelmänä saattaa auttaa mallia luokittelemaan paremmin kuvia, jotka ovat sumeita esimerkiksi sen vuoksi, että kamera tai kuvassa oleva esine on liikkeessä. (Shorten & Khoshgoftaar, 2019)

Data-augmentoinnin menetelmänä ydinsuodinten käytöstä tekee vähemmän hyödyllistä se, että se on saman kaltainen prosessi kuin se, joka tapahtuu konvoluutioneuroverkon sisällä. Ydinsuotimen käyttö voidaan toteuttaa paremmin niin että neuroverkossa on kerros sitä varten, kuin että ydinsuodinta sovelletaan erikseen opetusdataan data-augmentointimenetelmänä. (Shorten & Khoshgoftaar, 2019) Sen vuoksi ydinsuotimia ei käytetä tässä tutkielmassa data-augmentointimenetelmänä.

5. Tulokset

5.1. Perinteisten augmentointimenetelmien hyperparametrihaku

Jokaista perinteistä augmentointimenetelmää kohti kokeiltiin, parantaako se luokittelun tulosta. Joissakin menetelmissä käytettiin hyperparametreja, joita on suositeltu asiaa koskevissa tutkimuksissa. Toisissa paras hyperparametrien arvo selvitettiin kokeilemalla.

Kaikissa tapauksissa tehtiin datan augmentointi ja luokittelu luokalla 21, joka kattaa noin 48 % datasta.

Luokittelun tulokset erilaisia perinteisiä augmentointimenetelmiä käytettäessä on esitetty taulukossa 2. Värikylläisyyden, värisävyn ja kirkkauden arvot merkitsevät osuutta suurimmasta mahdollisesta vaihteluvälistä. Ne luvut on lihavoitu, jotka ovat suurempia kuin ilman augmentointia mitattu F1-mitta.

Taulukko 2. Perinteisten data-augmentointimenetelmien hyperparametrihaun tulokset.

Augmentointimenetelmä	F1-mitta
ei augmentointia	91,35 %
satunnaisen alueen poistaminen	91,31 %
saturaatio 0,5; värisävy 0,5	90,87 %
saturaatio 0,25; värisävy 0,5	91,02 %
saturaatio 0,1; värisävy 0,5	90,80 %
saturaatio 0,5; värisävy 0,25	**91,41 %**
saturaatio 0,25; värisävy 0,25	89,47 %
saturaatio 0,1; värisävy 0,25	91,20 %
saturaatio 0,5; värisävy 0,1	**91,84 %**
saturaatio 0,25; värisävy 0,1	**91,67 %**
saturaatio 0,1; värisävy 0,1	**91,70 %**
kohinan lisääminen	89,17 %
kirkkaus 1	91,04 %
kirkkaus 0,5	90,88 %
kirkkaus 0,25	**91,68 %**

Tulosten perusteella jotkin värisävyjen ja kylläisyyden muutoksen yhdistelmät ovat hyödyllisiä tällä datasetillä ja tässä luokittelutehtävässä. Lisäksi kirkkauden muutos on hyödyllinen augmentointimenetelmä. Satunnaisen alueen poistaminen ja kohinana lisääminen eivät ole näiden tulosten perusteella hyödyllisiä augmentointimenetelmiä. Kaikkien menetelmien F1-mittaa parantavat vaikutukset ovat kuitenkin hyvin pieniä.

Kokeiltiin, minkälainen tulos on, kun käytetään kaikkia luokittelutulosta parantavia menetelmiä yhdessä. Tähän yhdistelmään valittiin värimuunnokset niin että taulukossa 2 mainittu kylläisyysparametri oli 0,5, värisävy oli 0,1 ja kirkkaus 0,5. Kokeiltiin samaa myös luokilla 8 ja 16. Tulokset ovat taulukossa 3.

Taulukko 3. Luokittelutarkkuudet perinteisten data-augmentointimenetelmien yhdistelmällä.

Luokka ja augmentointi	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Luokka 21, ei augmentointia	91 %	94,69 %	88,24 %	91,35 %
Luokka 21, perinteisten augmentointimenetelmien yhdistelmä	91 %	91,08 %	**92,00 %**	**91,51 %**
Luokka 8, ei augmentointia	86 %	79,25 %	69,42 %	74,01 %
Luokka 8, augmentointimenetelmien yhdistelmä	86 %	**80,02 %**	68,48 %	73,80 %
Luokka 16, ei augmentointia	96 %	57,01 %	62,19 %	59,49 %
Luokka 16, augmentointimenetelmien yhdistelmä	96 %	55,50 %	**67,76 %**	**61,02 %**

Taulukon 3 mukaan augmentointimenetelmät parantavat tulosta luokalla 21 hieman myös yhdessä käytettyinä. Luokalla 8 ne taas eivät paranna tulosta F1-mitalla mitattuna. Luokalla 16 yhdistelmä taas paransi tulosta. Luokalla 21 yhdistelmä paransi tulosta hieman vähemmän kuin eräät taulukossa 2 mainitut menetelmät yksinään.

5.2. Vähemmistöluokan ylinäytteistys augmentoimalla

5.2.1 Ylinäytteistys GANilla, tyylinsiirrolla ja perinteisillä augmentointimenetelmillä generoiduilla kuvilla

Kokeiltiin vähemmistöluokan ylinäytteistystä data-augmentointimenetelmillä. Koe tehtiin kahdella vähemmistöluokalla: luokalla 8 ja luokalla 16.

Kumpaakin vähemmistöluokkaa laajennettiin siten, että generoitiin niin paljon vähemmistöluokkaan kuuluvaa dataa, että sitä oli saman verran kuin toiseen luokkaan kuuluvaa dataa. Toinen luokka tarkoittaa luokkaa, jonka kuvilla ei ole luokittelun kohteena olevaa attribuuttia.

Luokan 8 tapauksessa opetusdatassa on noin 39 000 siihen luokkaan kuuluvaa todellista kuvaa, joten toiseen luokkaan kuuluu noin 123 000 kuvaa. Sen vuoksi generoitiin noin 82 000 lisää luokkaan 8 kuuluvaa kuvaa, jotta molemmissa luokissa olisi saman verran kuvia. Samalla periaatteella luokan 16 luokittelua varten generoitiin noin 135 000 kuvaa.

GAN-menetelmällä voidaan generoida mielivaltaisen paljon uusia kuvia, jotka suurimmaksi osaksi eivät muistuta mitään yksittäistä alkuperäistä kuvaa. Tyylinsiirrolla generoitu kuva muistuttaa paljon sitä yksittäistä alkuperäistä kuvaa, johon on sovellettu satunnaista tyyliä. Siten GAN saattaa soveltua paremmin ylinäytteistykseen kuin tyylinsiirto. Vertailun vuoksi kokeiltiin kuitenkin myös tyylinsiirrolla ylinäytteistettyä dataa.

Lisäksi kokeiltiin augmentointia GANilla ja tyylinsiirrolla generoiduilla kuvilla, joihin on lisäksi sovellettu edellä mainittua perinteisten augmentointimenetelmien yhdistelmää. Sen kokeen tarkoituksena oli selvittää, onko GAN tai tyylinsiirto yhdistettynä perinteisiin menetelmiin augmentointimenetelmänä parempi kuin GAN tai tyylinsiirto sellaisenaan. Jos näin on, se voi oikeuttaa GANin tai tyylinsiirron käytön, vaikka GAN ja tyylinsiirto yksinään eivät parantaisi luokittelu tulosta verrattuna perinteisiin menetelmiin, tai olisivat perinteisiin menetelmiin nähden tarpeettoman monimutkaisia.

GANilla ja tyylinsiirrolla generoitujen kuvien lisäksi kokeiltiin tavallista ylinäytteistystä, jossa vähemmistöluokkaan kuuluvista kuvista tehdään identtisiä kopioita niin, että niitä on yhtä paljon kuin enemmistöluokkaan kuuluvia kuvia.

Luokittelun onnistumista mitattiin useilla mittareilla.

Luokan 8 osalta tulokset ovat taulukossa 4. Taulukossa on lihavoitu ne luvut, jotka ovat suurempia kuin samassa sarakkeessa oleva sellaisen luokittelun mittaustulos, jossa ei ole käytetty augmentointia. Samalla periaatteella on lihavoitu myös muut tässä luvussa olevien taulukkojen luvut.

Taulukko 4. CelebA-datasetin vähemmistöluokan 8 ylinäytteistys eri menetelmillä, ja niiden vaikutukset luokittelun onnistumiseen.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	86,76 %	79,25 %	69,42 %	74,01 %
Alkuperäiset + GANilla generoidut kuvat	85,37 %	68,65 %	**84,91 %**	**75,92 %**
Alkuperäiset + tyylinsiirrolla generoidut kuvat	83,81 %	67,68 %	**77,31 %**	72,18 %
Alkuperäiset + GANilla generoidut, joihin on kaikkiin sovellettu perinteisiä augmentointimenetelmiä	84,70 %	67,25 %	**85,10 %**	**75,13 %**
Alkuperäiset + tyylinsiirrolla generoidut, joihin on kaikkiin sovellettu perinteisiä augmentointimenetelmiä	83,89 %	67,86 %	**77,31 %**	72,28 %
Alkuperäiset + identtisiä kopioita	83,90 %	64,90 %	**88,79 %**	**74,98 %**

Luokan 16 osalta tulokset ovat taulukossa 5.

Taulukko 5. CelebA-datasetin luokan vähemmistöluokan 16 ylinäytteistys eri menetelmillä, ja niiden vaikutukset luokittelun onnistumiseen.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	96,12 %	57,01 %	62,19 %	59,49 %
Alkuperäiset + GANilla generoidut kuvat	90,46 %	30,69 %	**85,90 %**	45,22 %

Alkuperäiset + tyylinsiirrolla generoidut kuvat	90,57 %	30,82 %	**84,92 %**	45,23 %
Alkuperäiset + GANilla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	90,93 %	31,67 %	**84,48 %**	46,07 %
Alkuperäiset + tyylinsiirrolla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	92,76 %	36,64 %	**79,56 %**	50,17 %
Alkuperäinen + identtisiä kopioita	92,00 %	35,33 %	**90,93 %**	50,89 %

Tulosten perusteella ylinäytteistys kaikilla menetelmillä parantaa aina herkkyysarvoa ja heikentää tarkkuusarvoa sekä luokilla 8 että 16. Luokan 16 tapauksessa kaikki ylinäytteistysmenetelmät vähentävät F1-mitan arvoa. Luokan 8 tapauksessa generatiivisilla kilpailevilla verkostoilla tuotettujen kuvien käyttäminen datan augmentoinnissa parantaa F1-mitan arvoa, kun taas tyylinsiirron käyttäminen ei paranna sitä.

Luokan 8 tapauksessa identtisten kopioiden käyttäminen ylinäytteistyksessä parantaa F1-mitan arvoa, mutta luokan 16 tapauksessa ei.

Täsmällisyysluku on pienillä vähemmistöluokilla luultavasti kohtalaisen suuri, koska luokittelija saavuttaa korkean tarkkuuden ennustamalla, että suurin osa kuvista kuuluu enemmistöluokkaan.

Jos vähemmistöluokkaa augmentoidaan ylinäytteistystarkoituksessa perinteisillä augmentointimenetelmillä, kuten geometrisillä muunnoksilla, se saattaa aiheuttaa sen, että malli ylisovittuu siihen vähemmistöluokkaan. Se johtuu siitä, että vähemmistöluokan datassa olevat vääristymät ovat enemmän läsnä, kun sen luokan kuvista tehdään melkein identtisiä kopioita. (Shorten & Khoshgoftaar, 2019)

5.2.2 Alkuperäisen datan ja augmentointidatan suhde

Jos ylinäytteistys toteutetaan data-augmentointimenetelmillä, pienen vähemmistöluokan tapauksessa on niin, että vähemmistöluokan datapisteet ovat suurimmaksi osaksi augmentointimenetelmillä tuotettuja kuvia. Se voi olla ongelma, jos augmentointikuvat eivät ole niin hyvälaatuisia kuin alkuperäinen data, tai jos augmentointimenetelmä tuottaa kuvia, joiden ominaisuudet painottuvat eri tavalla kuin alkuperäisessä datassa. Sen vuoksi kokeiltiin ylinäytteistystä myös niin, että vähemmistöluokkaa täydennettiin eri määrillä augmentointidataa. Tätä kokeiltiin sekä generatiivisilla kilpailevilla verkostoilla että tyylinsiirrolla generoiduilla kuvilla.

Kokeiltiin augmentointia niin, että lisättiin koko alkuperäisen vähemmistöluokan dataan saman verran generoituja kuvia, kuin vähemmistöluokan alkuperäisessä datassa oli. Lisäksi kokeiltiin niin että lisättiin vähemmistöluokkaan generoituja kuvia 25 %, 50 %, 150 % ja 300 % alkuperäisten kuvien lukumäärän verran. Kokeiltiin myös alkuperäisten kuvien identtisillä kopioilla tehtyä ylinäytteistystä samoilla augmentointikuvien lukumäärillä. Vähemmistöluokkana käytettiin CelebA-datasetin luokkaa 8. Tulokset on esitetty taulukossa 6. Tyylinsiirtoa augmentointisuhteella 1:3 ei kokeiltu, koska muiden tulosten perusteella oletettiin, että se ei paranna luokittelun tarkkuutta.

Taulukko 6. Alkuperäisen datan datapisteiden ja augmentointidatan datapisteiden lukumäärän suhteen vaikutus luokittelutarkkuuteen CelebA-datasetin luokalla 8.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	87,14 %	78,23 %	72,96 %	75,50 %
GAN, 1:0,25	**87,33 %**	77,38 %	**75,40 %**	**76,38 %**
GAN, 1:0,5	85,71 %	69,38 %	**84,82 %**	**76,33 %**
GAN, 1:1	87,05 %	75,16 %	**78,16 %**	**76,63 %**
GAN, 1:1,5	86,59 %	73,01 %	**80,34 %**	**76,50 %**
GAN, 1:3	85,07 %	68,15 %	**84,53 %**	75,46 %
Tyylinsiirto, 1:0,25	**87,34 %**	76,19 %	**77,65 %**	**76,91 %**
Tyylinsiirto, 1:0,5	86,42 %	75,58 %	**73,88 %**	74,72 %
Tyylinsiirto, 1:1	85,36 %	70,55 %	**79,12 %**	74,59 %
Tyylinsiirto, 1:1,5	84,89 %	69,76 %	**78,31 %**	73,79 %
Identtiset kopiot, 1:0,25	87,10 %	74,23 %	**80,39 %**	**77,19 %**
Identtiset kopiot, 1:0,5	87,10 %	73,94 %	**81,06 %**	**77,34 %**
Identtiset kopiot, 1:1	**87,25 %**	74,44 %	**80,82 %**	**77,50 %**
Identtiset kopiot, 1:1,5	86,69 %	71,94 %	**83,60 %**	**77,34 %**
Identtiset kopiot, 1:3	84,05 %	64,67 %	**91,02 %**	**75,61 %**

Tulosten perusteella paras suhde alkuperäisen ja GANilla generoidun datan välillä CelebA-datasetin luokan 8 tapauksessa on noin 1:1. Tyylinsiirron tapauksessa parannus tarkkuuteen saavutettiin vain suhteella 1:0,25. Luvun 5.3 tulosten perusteella tyylinsiirto sopii paremmin augmentointimenetelmäksi muiden CelebA-datasetin luokkien kuin luokan 8 tapauksessa, jolla tämä koe tehtiin. Tavallinen ylinäytteistys ilman generatiivisilla kilpailevilla verkostoilla tai tyylinsiirrolla tehtyä data-augmentointia parantaa tarkkuutta eniten.

5.2.3 Datasetin täydentäminen CycleGANilla generoiduilla kuvilla

Kokeiltiin data-augmentointia sellaisilla kuvilla, jotka on generoitu CycleGANilla toisen luokan pohjalta. Opetettiin CycleGAN muuttamaan enemmistöluokan kuvia vähemmistöluokan kuviksi, ja käytettiin niitä vähemmistöluokan täydentämiseen.

Muissa tässä tutkielmassa tehdyissä kokeissa augmentointidataa tehtiin saman luokan pohjalta. Ensin mainittu menetelmä on parempi sillä tavalla, että augmentointi positiivisen luokan data sisältää uutta tietoa alkuperäiseen dataan verrattuna.

Muualla tässä tutkielmassa käytetty GAN myös vaatii suuren määrän dataa laadukkaiden kuvien tuottamiseen. Se vaatii kymmeniä tuhansia kuvia tuottaakseen dataa, joka näyttää yhtä laadukkaalta kuin alkuperäinen data. CycleGANille riittää taas esimerkiksi tuhat kuvaa. Toisaalta CycleGANin opettaminen vaatii huomattavasti enemmän laskentaresursseja. Tätä menetelmää kokeiltiin, koska sen arveltiin suoriutuvat pelkän vähemmistöluokan datalla opetettua DCGANia paremmin ainakin silloin, kun käytössä on vain vähän vähemmistöluokan dataa.

Koska CycleGANin opettaminen vaatii paljon laskentaresursseja, koe tehtiin vain pienellä osalla CelebA-datasetin datasta.

Opetettiin CycleGAN syöttämällä sille 1000 positiivisen luokan kuvaa ja 1000 negatiivisen luokan kuvaa. Positiivisena luokkana käytettiin CelebA-datasetin luokkaa 8, ja negatiivinen luokka oli siten ne kuvat, jotka eivät kuuluneet luokkaan 8. Lisäksi tehtiin koe, jossa positiivisena luokkana oli luokka 35. Siten CycleGAN oppi muuttamaan sille syötetyt negatiivisen luokan kuvat positiivisen luokan kuviksi.

Vaikka käytettiin vain osaa CelebA-datasetin kuvista, tarkoitus oli simuloida tilannetta, jossa on vähemmistöluokka, jota kannattaa mahdollisesti täydentää CycleGANilla generoidulla datalla. Opetettiin luokittaja 1000 positiivisen luokan kuvalla ja 2000 negatiivisen luokan kuvalla, jolloin positiivinen luokka oli vähemmistöluokka. Käytettiin samaa dataa, jolla opetettiin CycleGAN. Näin mitattiin luokittelun tarkkuus ilman data-augmentointia.

Sitten opetettiin luokittaja niin, että positiivista luokkaa täydennettiin 1000 CycleGANilla generoidulla kuvalla. Ne kuvat tehtiin 1000 negatiivisen luokan kuvan pohjalta. Ne negatiivisen luokan kuvat ovat samoja, joita myös käytettiin luokittajan opetuksessa. Tässä kokeessa oli siis 1000 alkuperäistä ja 1000 generoitua positiivisen luokan kuva, ja 2000 alkuperäistä negatiivisen luokan kuvaa. Näin mitattiin luokittelun tarkkuus, kun ylinäytteistyksessä käytettiin data-augmentointia CycleGANilla.

Lisäksi tehtiin koe ylinäytteistyksellä niin, että ylinäytteistyksessä käytettiin positiivisen luokan kuvien kopioita.

Vertailun vuoksi tehtiin koe myös niin, että augmentointimenetelmänä käytettiin tyylinsiirtoa. Generoitiin tyylinsiirrolla 1000 positiivisen luokan kuvaa alkuperäisten positiivisen luokan kuvien pohjalta.

Tulokset ovat taulukoissa 7 ja 8.

Taulukko 7. Data-augmentointi CycleGANilla verrattuna muihin ylinäytteistystapoihin CelebA-datasetin luokalla 8.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	82,73 %	66,46 %	73,52 %	69,81 %
Ylinäytteistys CycleGANilla generoiduilla kuvilla	79,51 %	58,72 %	**82,74 %**	68,69 %
Ylinäytteistys tyylinsiirrolla generoiduilla kuvilla	79,26 %	59,18 %	**76,21 %**	66,62 %
Ylinäytteistys identtisillä kuvilla	81,60 %	63,12 %	**77,59 %**	69,61 %

Taulukko 8. Data-augmentointi CycleGANilla verrattuna muihin ylinäytteistystapoihin CelebA-datasetin luokalla 35.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	94,99 %	45,02 %	86,77 %	59,28 %
Ylinäytteistys CycleGANilla generoiduilla kuvilla	91,67 %	32,38 %	**90,11 %**	47,64 %
Ylinäytteistys tyylinsiirrolla generoiduilla kuvilla	91,77 %	32,78 %	**91,18 %**	48,22 %
Ylinäytteistys identtisillä kuvilla	93,93 %	39,92 %	**87,84 %**	54,90 %

Datan augmentointi generatiivisilla kilpailevilla verkostoilla tai tyylinsiirrolla luoduilla kuvilla ei parantanut tuloksia kummankaan luokan tapauksessa. Ylinäytteistys identtisillä kuvilla ei myöskään parantanut luokittelun tuloksia.

Jos kokeet tehtäisiin eri määrällä dataa, tai augmentointidatan suhdetta alkuperäiseen dataan muutettaisiin, tulokset voisivat olla erilaisia.

5.3. Koko datasetin täydentäminen

Vähemmistöluokan ylinäytteistyksen lisäksi kokeiltiin GAN- ja tyylinsiirtodata-augmentointimenetelmien vaikutusta luokittelun onnistumiseen siten, että lisättiin niillä menetelmillä generoitua dataa molempiin luokkiin. Luokittajan opetusdatassa käytettiin kaikkea todellista opetusdataa ja sen lisäksi saman verran generoitua dataa. Generoitua dataa oli siis alkuperäiseen dataan suhteessa 1:1. Luokittelussa oli kerrallaan aina kaksi luokkaa, ja kumpaankin luokkaan kuuluvaa dataa generoitiin sen verran että luokkien datapisteiden lukumäärien välinen suhde säilyi.

Tulokset on esitetty taulukoissa 9, 10 ja 11.

Taulukko 9. CelebA-datasetin luokan 21 luokittelun tulokset eri data-augmentointimenetelmillä, kun koko datasettiin on sovellettu data-augmentointia.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	91,73 %	94,69 %	88,24 %	91,35 %
Alkuperäiset + GANilla generoidut kuvat	90,56 %	88,52 %	**93,00 %**	90,70 %
Alkuperäiset + tyylinsiirrolla generoidut kuvat	**91,79 %**	91,37 %	**92,12 %**	**91,74 %**
Alkuperäiset + GANilla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	90,73 %	91,67 %	**89,40 %**	90,52 %
Alkuperäiset + tyylinsiirrolla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	**92,13 %**	90,81 %	**93,56 %**	**92,17 %**

Taulukko 10. CelebA-datasetin luokan 8 luokittelun tulokset eri data-augmentointimenetelmillä, kun koko datasettiin on sovellettu data-augmentointia.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	86,76 %	79,25 %	69,42 %	74,01 %
Alkuperäiset + GANilla generoidut kuvat	83,69 %	68,17 %	**74,97 %**	71,41 %
Alkuperäiset + tyylinsiirrolla generoidut kuvat	86,30 %	79,13 %	67,34 %	72,76 %
Alkuperäiset + GANilla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	84,03 %	70,40 %	**71,12 %**	70,76 %
Alkuperäiset + tyylinsiirrolla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	85,39 %	**79,51 %**	62,26 %	69,84 %

Taulukko 11. CelebA-datasetin luokan 16 luokittelun tulokset eri data-augmentointimenetelmillä, kun koko datasettiin on sovellettu data-augmentointia.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	96,12 %	57,01 %	62,19 %	59,49 %
Alkuperäiset + GANilla generoidut kuvat	**96,48 %**	**64,13 %**	52,57 %	57,78 %
Alkuperäiset + tyylinsiirrolla generoidut kuvat	**96,43 %**	**60,90 %**	61,97 %	**61,43 %**
Alkuperäiset + GANilla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	**96,33 %**	**60,34 %**	58,03 %	59,16 %
Alkuperäiset + tyylinsiirrolla generoidut, joihin kaikkiin on sovellettu perinteisiä augmentointimenetelmiä	**96,28 %**	**58,71 %**	**63,72 %**	**61,11 %**

Tulosten perusteella koko datasetin laajentaminen generatiivisilla kilpailevilla verkostoilla generoidulla datalla ei johdonmukaisesti paranna tai huononna luokittelun tarkkuutta. Koko datasetin laajentaminen generoidulla datalla ei siis vaikuta olevan hyödyllistä, vaikka vähemmistöluokan ylinäytteistys generoidulla datalla vaikuttaisi olevan hyödyllistä. Tutkimuksessa (Bowles ym., 2018) havaittiin että augmentointidatan lisääminen alkuperäisen datan joukkoon parantaa tuloksia pääsääntöisesti sitä enemmän, mitä vähemmän alkuperäistä dataa on saatavilla. Nämä mittaustulokset ovat linjassa sen kanssa siinä mielessä, että alkuperäistä dataa on saatavilla varsin paljon, koska opetusdatasetin koko on 160 000 kuvaa, kun taas vähemmistöluokissa sitä on vähemmän.

Kun data-augmentoinnissa käytettiin tyylinsiirtoa, tulokset paranivat hieman luokkien 21 ja 16 tapauksessa.

5.4. Alkuperäisen datan korvaava data-augmentointi

GANia ei voi käyttää tyylinsiirrolle suoraan vertailukelpoisena augmentointimenetelmänä silloin, kun korvataan alkuperäistä dataa, koska GAN generoi

kokonaan uuden kuvan, eikä muuta alkuperäistä kuvaa. Kokeiltiin koko datasetin korvaamista GANilla generoiduilla kuvilla. Tyylinsiirron tapauksessa taas kokeiltiin tyylinsiirron soveltamista osaan kuvista, samalla tavalla kuin perinteisiä augmentointimenetelmiä usein sovelletaan.

On mahdollista, että GAN-menetelmää ei kannata käyttää korvaamaan alkuperäistä dataa, jos tavoitteena on parantaa luokittelun tarkkuutta. Se johtuu siitä, että GANilla on usein vaikea saada tuotettua kuvia, jotka ovat yhtä laadukkaita alkuperäiset kuvat. Näin on varsinkin sellaisissa tilanteissa, joissa opetusdataa GANille on tarjolla vähän. Ne ovat usein samoja tilanteita, joissa datan augmentoinnista olisi eniten hyötyä, koska lisähyöty on luultavasti pienempi, jos luokittelijan opetusdataa on jo paljon saatavilla.

Aiemmassa luvussa mainittiin, että on ehdotettu, että GAN-menetelmää datan augmentointiin voisi käyttää tilanteissa, joissa syystä tai toisesta ei haluta käyttää alkuperäistä dataa. Syynä siihen voi olla esimerkiksi tietosuoja, jos data on henkilöön liittyvää dataa. Sen vuoksi tässä tutkielmassa kokeiltiin alkuperäisen datan korvaavaa data-augmentointia GAN-augmentointimenetelmällä.

GANin tapauksessa generoitiin saman verran kuvia kuin kussakin luokassa oli kuvia. Sitten alkuperäiset kuvat korvattiin niillä.

Luokalla 21 luokittelutarkkuudeksi saatiin F1-mitalla mitattuna noin 59,07 %, kun taas ilman augmentointimenetelmiä luokittelun tarkkuus F1-mitalla mitattuna oli 91,35 %. Luokalla 8 vastaavat luvut olivat 43,79 % ja 74,01 %. Näiden tulosten perusteella pääteltiin, että tässä tutkielmassa käytetyt generatiiviset kilpailevat verkostot ja niiden opetusmenetelmät eivät voi tuottaa tarpeeksi laadukkaita kuvia sellaista data-augmentointia varten, jossa augmentoitu data korvaa alkuperäisen datan.

Tyylinsiirron tapauksessa korvattiin puolet datasetistä kuvilla, joihin oli sovellettu tyylinsiirtoa. Luokittajan opetuksen aikana jokaisen epookin kohdalla valittiin satunnaisesti, mihin kuviin sovelletaan tyylinsiirtoa. Siten tämä datan augmentoinnin menetelmä on saman kaltainen kuin jos lisättäisiin datasettiin alkuperäisten kuvien lisäksi tyylinsiirrolla generoituja kuvia.

Tulokset on esitetty taulukoissa 12–15.

Taulukko 12. Luokittelu CelebA-datasetin luokilla 21 ja 8, kun koko datasetti on korvattu data-augmentointimenetelmillä generoiduilla kuvilla.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Luokka 21, ei augmentointia	91,73 %	94,69 %	88,24 %	91,35 %
Luokka 21, vain GANilla generoidut kuvat	64,21 %	68,09 %	52,16 %	59,07 %
Luokka 8, ei augmentointia	86,76 %	79,25 %	69,42 %	74,01 %
Luokka 8, vain GANilla generoidut kuvat	69,25 %	43,48 %	44,10 %	43,79 %

Taulukko 13. Luokittelu CelebA-datasetin luokalla 21, kun osa datasetistä on korvattu vuorotellen tyylinsiirrolla generoidulla datalla.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	91,73 %	94,69 %	88,24 %	91,35 %
Tyylinsiirto	**91,82 %**	92,31 %	**91,07 %**	**91,68 %**
Tyylinsiirto ja perinteiset augmentointimenetelmät	**91,78 %**	93,52 %	**89,61 %**	**91,52 %**

Taulukko 14. Luokittelu CelebA-datasetin luokalla 8, kun osa datasetistä on korvattu vuorotellen tyylinsiirrolla generoidulla datalla.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	86,76 %	79,25 %	69,42 %	74,01 %
Tyylinsiirto	85,64 %	78,03 %	65,58 %	71,27 %
Tyylinsiirto ja perinteiset augmentointimenetelmät	85,23 %	78,17 %	63,28 %	69,94 %

Taulukko 15. Luokittelu CelebA-datasetin luokalla 16, kun osa datasetistä on korvattu vuorotellen tyylinsiirrolla generoidulla datalla.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	96,12 %	57,01 %	62,19 %	59,49 %
Tyylinsiirto	**96,49 %**	**63,00 %**	56,94 %	**59,82 %**
Tyylinsiirto ja perinteiset augmentointimenetelmät	**96,42 %**	**60,82 %**	61,75 %	**61,28 %**

Tulosten perusteella vaikutta siltä, että pelkkiä generatiivisilla kilpailevilla verkostoilla tuotettuja kuvia käyttämällä ei saada kelvollisia luokittelutuloksia.

Wei ym. (2020) tutkivat myös luokittelijan opettamista pelkästään GANilla generoiduilla kuvilla. Luokittelutarkkuus oli huono, vaikka ihmisasiantuntijat arvioivat generoidut kuvat laadukkaiksi, ja vaikka GANilla generoitujen kuvien käyttö yhdessä alkuperäisen datan kanssa paransi tuloksia.

Huonot tulokset muihin kokeisiin verrattuna voivat johtua osittain siitä, että alkuperäisen data korvaaminen tehtiin samalla menetelmällä kuin perinteisten augmentointimenetelmien käyttö, eli generoitiin joka epookilla uudet kuvat.

5.5. EuroSAT-datasetin kokeet

Data-augmentointimenetelmien vaikutusta selvitettiin myös EuroSAT-datasetillä. Sillä tehtiin suppeampi määrä kokeita.

EuroSAT-datasetillä kokeiltiin vähemmistöluokan ylinäytteistystä. Vähemmistöluokaksi valittiin kerrallaan yksi luokista, niin että muut luokat muodostivat toisen luokan. Vähemmistöluokkaa täydennettiin siten, että generoitiin uusia kuvia saman luokan kuvien pohjalta. Lisäksi kokeiltiin tavallista ylinäytteistystä, eli datasettiä täydennettiin vähemmistöluokan kopioilla.

EuroSAT-datasetillä luokittajan opetusta tehtiin kymmenen epookin ajan, koska sen jälkeen tulokset eivät usein parantuneet augmentoimattomalla datalla. Tulokseksi valittiin sen epookin tulos, joka oli paras F1-mitan perusteella. Viimeistä tulosta ei käytetty, jos se ei ollut paras, koska on mahdollista, että malli ylisovittuu opetuksen aikana ja tulokset heikkenevät sen vuoksi opetuksen edetessä.

Kokeet tehtiin Forest- ja River-luokilla, joiden kuvissa on metsää ja jokea. Forest-luokassa olennainen ominaisuus on luultavasti tekstuuri, kun taas River-luokassa on myös muotoja eli joen rantaviivaa.

Tulokset on esitetty taulukoissa 16 ja 17.

Taulukko 16. EuroSAT-datasetin Forest-luokan luokittelun tulokset.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	97,58 %	88,73 %	90,50 %	89,60 %
Alkuperäiset + GANilla generoidut kuvat	79,02 %	34,01 %	87,00 %	48,90 %
Alkuperäiset + tyylinsiirrolla generoidut kuvat	59,90 %	21,37 %	**92,33 %**	34,70 %
Yksinkertainen ylinäytteistys	93,75 %	66,23 %	**93,50 %**	77,54 %

Taulukko 17. EuroSAT-datasetin River-luokan luokittelun tulokset.

	Täsmällisyys	Tarkkuus	Herkkyys	F1-mitta
Ei augmentointia	95,02 %	77,83 %	67,40 %	72,24 %
Alkuperäiset + GANilla generoidut kuvat	91,69 %	54,76 %	**78,20 %**	64,42 %
Alkuperäiset + tyylinsiirrolla generoidut kuvat	51,65 %	16,00 %	**94,80 %**	27,38 %
Yksinkertainen ylinäytteistys	90,67 %	50,85 %	**89,40 %**	64,83 %

Tulosten perusteella generatiiviset kilpailevat verkostot ja tyylinsiirto eivät ole hyödyllisiä augmentointimenetelmiä EuroSAT-datasetin laajentamiseen. Generatiivisilla kilpailevilla verkostoilla tuotettujen kuvien käyttäminen augmentoinnissa heikentää

luokittelun onnistumista merkittävästi, ja tyylinsiirron käyttäminen heikentää sitä todella paljon.

Tyylinsiirto augmentointimenetelmänä suoriutui huonosti EuroSAT-datan augmentointimenetelmänä mahdollisesti osittain siksi, koska tyylinsiirto muuttaa kuvien tekstuureja ja värejä, ja ne ovat luultavasti olennaisia kuvien tunnistamisessa.

Tulosten perusteella myöskään tavallinen ylinäytteistys eli identtisten kopioiden lisääminen ei paranna luokittelutulosta. Se kuitenkin heikentää luokittelun onnistumista vähemmän kuin datan augmentointi generatiivisilla kilpailevilla verkostoilla tai tyylinsiirrolla.

Datasetin laajentaminen GANeilla generoiduilla kuvilla ei luultavasti parantanut luokittelun tuloksia sen vuoksi, että generoidut kuvat olivat heikkolaatuisia. Esimerkiksi CelebA-datasetin pohjalta GANilla generoidut kuvat olivat laadukkaampia, koska GAN voitiin opettaa suuremmalla määrällä dataa, koska CelebA-datasetti sisältää enemmän kuvia kuin EuroSAT.

6. Tutkielman soveltuvuus ja rajoitukset

6.1. Datan ja konenäkötehtävien rajaus

Tutkielma koskee lähinnä CelebA-datasetin kuvia, ja tulokset eivät välttämättä yleisty hyvin muun tyyppiseen kuvadataan, tai muihin kasvokuvadatasetteihin.

Erityisesti tyylinsiirron tarkastelemiseen data-augmentointimenetelmänä olisi sopinut joku muu kuin EuroSAT-datasetti, koska EuroSAT-datasetin kuvissa tekstuurit ovat luultavasti olennaisia, ja tyylinsiirto muuttaa niitä.

Tässä tutkielmassa vertailtiin data-augmentointimenetelmien soveltuvuutta ainoastaan kuvien luokittelussa. Tulokset eivät luultavasti yleisty muihin konenäkötehtäviin, joita ovat esimerkiksi esineen tunnistaminen kuvassa, kuvan jakaminen osiin ja syvyyden arvioiminen kuvassa. Kuten aikaisemmassa luvussa mainittiin, erityisesti tyylinsiirto saattaa soveltua paremmin augmentointimenetelmäksi syvyyden arvioinnissa verrattuna muihin augmentointimenetelmiin.

6.2. Datan generoiminen data-augmentointimenetelmillä

DCGANin opetuksessa virhearvot vaihtelivat paljon opetuksen aikana yhden epookin sisällä. Olisi voitu saavuttaa parempi generaattorimalli lopettamalla opetus sellaisessa vaiheessa, jossa virhearvot olivat hyvät.

Perinteisten data-augmentointimenetelmien hyperparametrihaku tehtiin vain luokalla 21. Siten tässä tutkielmassa käytetyt perinteiset data-augmentointimenetelmät ja niiden parametrit eivät välttämättä sovellu hyvin muilla CelebA-datasetin luokille. Ainakin satunnaisen alueen peittäminen ja värisävyjen ja kirkkauden muunnokset ovat luultavasti sellaisia data-augmentointimenetelmiä, jotka sopivat paremmin joihinkin luokkiin kuin toisiin.

CelebA-datasetin pienen vähemmistöluokan 16 tapauksissa käytettiin ylinäytteistyksessä suhteellisesti paljon augmentointidataa verrattuna alkuperäisten datapisteiden määrään, mikä ei luultavasti ollut optimaalista tulosten perusteella, jotka on esitetty aliluvussa 5.2.2. Siten luokan 16 ylinäytteistyksen tulokset eivät kerro kovin hyvin, kuinka paljon hyötyä käytetyistä data-augmentoinnin menetelmistä voi olla sen tapauksessa. Muidenkin luokkien kuin luokan 16 tapauksessa augmentointidatan ja alkuperäisen datan suhteella

lienee merkittävä vaikutus, ja kokeiden tekeminen vain yhdellä suhteella antaa puutteellisen kuvan augmentointimenetelmien hyödyllisyydestä.

Luokittelussa kuvat kutistettiin ennen luokittelijalle syöttämistä niin, että niiden kooksi tulo 64x64. Se saattoi erityisesti luokan 16 tapauksessa häivyttää olennaisia kohtia kuvista.

CycleGANin tuottamien kuvien laatua olisi voitu parantaa siten, että olisi valittu niistä vain parhaat kuvat. Parhaat kuvat olisi voinut valita seuraavalla menetelmällä: Ensin opetetaan luokittaja todellisella opetusdatalla luokittelemaan kuvat siihen luokkaan, johon kuuluvia kuvia halutaan generoida. Sitten luokitellaan generoidut kuvat, ja valitaan niistä ainoastaan ne, joille luokittelija antaa määrätyn kynnysarvon ylittävän todennäköisyyden, että ne kuuluvat kyseiseen luokkaan. Näin tehtiin tutkimuksessa (Wei ym., 2020).

Kuvien generoinnissa ja luokittelussa käytettävien neuroverkkojen rakenne ja hyperparametrit eivät välttämättä ole optimaalisia, koska ei selvitetty kokeellisesti, mikä rakenne ja mitkä hyperparametrit sopivat parhaiten käytetylle datalle.

6.3. Luokittelun toteutus

Tässä tutkielmassa tutkittiin ainoastaan tapauksia, joissa data luokitellaan kahteen luokkaan. Olisi voinut tutkia myös tapauksia, joissa luokkia on enemmän. Toisaalta sellaiset luokittelutehtävät, joissa luokkia on monta, voidaan tyypillisesti jakaa sellaisiksi luokittelutehtäviksi, joissa luokkia on kaksi, joten toisaalta siksi on perusteltua keskittyä vain tapauksiin, joissa on kaksi luokkaa (Leevy ym., 2018).

Luokittelun tarkkuuteen saattaa vaikuttaa se, että CelebA-datasetissä data jaettiin kahteen osaan yhden ominaisuuden perusteella, kuten sen perusteella onko kuvassa olevalla henkilöllä leukaparta. Tämä saattaa aiheuttaa sen, että se luokka, jossa ominaisuutta ei ole läsnä, sisältää enemmän sisäistä vaihtelua kuin luokka, jossa ominaisuus on läsnä. Jos data olisi jaettu moneen luokkaan, luokkien sisäinen vaihtelu olisi luultavasti ollut pienempi. Moneen luokkaan luokittelua ei voi helposti kokeilla CelebA-datasetillä, koska siinä monet luokat ovat keskenään päällekkäisiä.

Näillä koejärjestelyillä on vaikea erottaa toisistaan sitä, johtuvatko erot luokittelun tuloksissa luokkien ominaispiirteistä vai luokan datapisteiden määrästä. Se johtuu siitä,

että datapisteiden määrä on erilainen eri luokilla, ja käytettiin kaikkia datapisteitä sen sijaan että olisi jätetty osa datasta pois siten, että kaikilla luokilla olisi ollut saman verran dataa. Toisaalta datamäärän rajoittaminen olisi vaikuttanut DCGANilla tuotettujen kuvien laatuun.

Joissakin tapauksissa luokittelun tulokset olisivat voineet olla parempia, jos opetusta olisi jatkettu pidempään. Opetusta olisi voinut jatkaa siihen asti, että validointidatasetillä mitattu tarkkuus olisi alkanut johdonmukaisesti laskea samalla kun opetussetillä mitattu tarkkuus olisi johdonmukaisesti noussut, jolloin olisi voinut tulkita, että malli on ylisovittunut. Sen jälkeen olisi voinut valita opetuksen kestoksi sen epookkien määrän, jolla saavutettiin paras validointidatasetin luokittelutulos. Sen jälkeen olisi voitu mitata sillä epookkien määrällä opetetun mallin tarkkuus testidatasetillä. Nyt luokittelijan opetus lopetettiin aina sen jälkeen, kun oli kulunut vakiomäärä epookkeja. Toisaalta jos joku augmentointimenetelmä parantaa parasta tulosta, joka saavutetaan ennen ylisovitusta, on mahdollista, että se augmentointimenetelmä yleensä myös parantaa luokittelutulosta jo opetuksen aikaisemmassa vaiheessa.

Luokittajan opetuksessa sekä augmentointimenetelmien soveltamisessa satunnaisuudella on rooli, ja sattuma voi selittää osan niistä tuloksista, joissa mitatut erot luokittelutarkkuuksien välillä olivat pienet. Luokittelun tarkkuudessa on jonkin verran vaihtelua, vaikka data ja parametrit ovat samat.

Luokittelun tarkkuus ja herkkyys vaihtelivat monessa tapauksessa epookista toiseen, joten niiden lopullisesta arvosta ei välttämättä voi päätellä paljon siitä, kuinka paljon käytetty augmentointimenetelmä parantaa niitä arvoja. Koska tarkkuusarvon parantaminen tyypillisesti heikentää herkkyysarvoa ja päinvastoin, F1-mitta ei kuitenkaan vaihdellut paljon epookista toiseen.

7. Yhteenveto

Tämän tutkielman alussa mainitut tutkimuskysymykset ovat: parantavatko eri menetelmät luokittelun tuloksia ja kuinka paljon, kun tarkastellaan ylinäytteistystä, koko datasetin täydentämistä, ja koko datasetin korvaamista data-augmentointimenetelmillä generoiduilla kuvilla?

Tämän tutkielman kokeiden tulosten perusteella ylinäytteistys GANilla parantaa F1-mitalla mitattuna joissakin tapauksissa luokittelun tuloksia enemmän kuin tavallinen ylinäytteistys datapisteiden kopioilla ilman data-augmentointia. Perinteisten data-augmentointimenetelmien käyttäminen GANin kanssa ei paranna tuloksia verrattuna siihen, että käytettiin pelkkää GANia. Tyylinsiirron käyttäminen ei tyypillisesti paranna tuloksia ylinäytteistyksessä.

Tämä on linjassa Bowlesin ym. (2018) tutkimuksen kanssa, jossa myös havaittiin, että GANilla luoduilla kuvilla augmentoitu data parantaa konenäkötehtävän tuloksia. Toisaalta kyseisen tutkimuksen mukaan synteettisen datan käyttö ei koskaan huononna tuloksia, kun taas tässä tutkielmassa se huononsi joissakin tapauksissa. Ero johtunee ainakin osittain siitä, että Bowles ym. käyttivät vähemmän synteettistä dataa suhteessa alkuperäiseen dataan kuin tässä tutkielmassa. Kyseisessä tutkimuksessa myös todettiin, että GANin ja perinteisten augmentointimenetelmien käyttö yhdessä tuottaa parempia tuloksia kuin kummankin käyttö yksinään.

CycleGANilla tehty vähemmistöluokan ylinäytteistys ei parantanut tuloksia tässä tutkielmassa tehdyissä kokeissa. Sitä vastoin esimerkiksi Wein yms. (2020) tutkimuksessa se paransi luokittelun tuloksia, kun sitä käytettiin vähemmistöluokan täydentämiseen.

Tämän tutkielman tulosten mukaan koko datasetin täydennyksessä GAN ei paranna luokittelun tuloksia. Tyylinsiirto parantaa joissakin tapauksissa tuloksia. Perinteisten data-augmentointimenetelmien käyttäminen tyylinsiirron rinnalla ei juurikaan vaikuta tuloksiin verrattuna siihen, että käytettäisiin vain tyylinsiirtoa.

Joissakin tutkimuksissa (Wang ja Perez, 2017; Zheng ym., 2019) tyylinsiirrolla tehty augmentointi paransi myös tuloksia vähän, samoin kuin tässä tutkielmassa. Niissä havaittiin myös, että perinteiset augmentointimenetelmät yhdessä tyylinsiirron kanssa tuottivat parempia tuloksia kuin pelkästään tyylinsiirron käyttäminen. GANin osalta eräissä tutkimuksissa (Bowles ym., 2018; Madani ym., 2018) todettiin, että GANien käyttäminen augmentointimenetelmänä oli hyödyllistä silloin, kun täydennettiin datasetin kaikkia luokkia, kun taas tässä tutkielmassa tehdyissä kokeissa se ei ollut.

Pienellä vähemmistöluokilla GANilla luodut kuvat eivät parantaneet luokittelun onnistumista ylinäytteistyksessä käytettynä. Se johtunee siitä, että tässä tutkielmassa käytetty GAN tarvitsee enemmän opetusdataa, jotta se voisi tuottaa tyydyttävän laatuisia kuvia. Tämä ongelma on sekä CelebA-datasetin että EuroSAT-datasetin kuvilla.

Tässä tutkielmassa koko datan korvaaminen GAN:illa generoidulla datalla ei tuottanut tyydyttäviä tuloksia.

Tämän tutkielman tulosten perusteella perinteiset augmentointimenetelmät eivät juuri paranna luokittelun tarkkuutta CelebA-datasetillä.

Ylinäytteistys parantaa kaikissa tapauksissa herkkyysarvoa, ja heikentää tarkkuusarvoa. Näin on sekä generatiivisilla kilpailevilla verkostoilla ja tyylinsiirrolla tuotettuja kuvia käytettäessä että alkuperäisten kuvien identtisiä kopioita käytettäessä. Ylinäytteistystä voidaan siis käyttää sellaisissa tilanteissa, joissa herkkyysarvoa pidetään jostakin syystä erityisen tärkeänä.

EuroSAT-datasetillä luokittelun tulokset eivät parantuneet, kun generatiivisilla kilpailevilla verkostoilla tai tyylinsiirrolla täydennettiin vähemmistöluokkaa. Tulokset eivät parantuneet myöskään, kun käytettiin sellaista ylinäytteistystä, jossa vähemmistöluokkaa täydennettiin vähemmistöluokan kuvien kopioilla.

Kaikki mitatut parannukset olivat hyvin pieniä, joten niistä ei välttämättä voi vetää johtopäätöksiä tässä tutkielmassa käytettyjen data-augmentointimenetelmien hyödyllisyydestä.

Yksi mahdollinen jatkotutkimuskohde on, että tehtäisiin samat kokeet sellaisilla generatiivisilla kilpailevilla verkostoilla tai niiden opetusmenetelmällä, jotka pystyvät tuottamaan hyvälaatuisia kuvia, ja jotka tarvitsevat vähemmän opetusdataa.

Samanlaisia kokeita voisi tehdä myös erilaisilla alkuperäisen datan ja augmentointia varten generoidun datan lukumäärillä.

Lähteet

Connor Shorten ja Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. 2019. Journal of Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

Philip T. Jackson, Amir Atapour-Abarghouei, Stephen Bonner, Toby P. Breckon ja Boguslaw Obara. Style Augmentation: Data Augmentation via Style Randomization. 2018. arXiv:1809.05375v2 https://arxiv.org/abs/1809.05375v2

Jason Wang ja Luis Perez. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 2017. arXiv:1712.04621 [cs.CV] https://arxiv.org/abs/1712.04621v1

Leon A. Gatys, Alexander S. Ecker ja Matthias Bethge. A Neural Algorithm of Artistic Style. 2015. arXiv:1508.06576 [cs.CV]

Christopher Bowles, Liang Chen, Ricardo Guerrero, Paul Bentley, Roger Gunn, Alexander Hammers, David Alexander Dickie, Maria Valdes Hernandez, Joanna Wardlaw, and Daniel Rueckert. GAN augmentation: augmenting training data using generative adversarial networks. 2018. arXiv:1810.10863 [cs.CV]

Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang. Deep Learning Face Attributes in the Wild. 2015. Proceedings of International Conference on Computer Vision (ICCV). https://arxiv.org/abs/1411.7766

Sergey Ioffe ja Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015. arXiv:1502.03167v3 [cs.LG] https://arxiv.org/abs/1502.03167

Mehdi Mirza ja Simon Osindero. Conditional Generative Adversarial Nets. 2015. arXiv:1411.1784v1 https://arxiv.org/abs/1411.1784

Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li ja Yi Yang. Random Erasing Data Augmentation. 2017. arXiv:1708.04896v2 [cs.CV] https://arxiv.org/abs/1708.04896

Francisco J. Moreno-Barea, Fiammetta Strazzera, José M. Jerez ja D. Urda, Leonardo Franco. Forward Noise Adjustment Scheme for Data Augmentation. 2018. IEEE Symposium Series on Computational Intelligence (SSCI)

Luke Taylor ja Geoff Nitschke. Improving Deep Learning using Generic Data Augmentation. 2017. arXiv:1708.06020v1 https://arxiv.org/abs/1708.06020

Yi Liu, Jialiang Peng, James J.Q Yu ja Yi Wu. PPGAN: Privacy-preserving Generative Adversarial Network. 2019. arXiv:1910.02007 [cs.LG] https://arxiv.org/abs/1910.02007v1

Chuan Ma, Jun Li, Ming Ding, Bo Liu, Kang Wei, Jian Weng ja H. Vincent Poor. RDP-GAN: A Rényi-Differential Privacy based Generative Adversarial Network. 2020. arXiv:2007.02056 [cs.LG] https://arxiv.org/abs/2007.02056v1

Alex Krizhevsky, Ilya Sutskever ja Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012. Advances in Neural Information Processing Systems 25 (NIPS 2012)

Fabio Perez, Cristina Vasconcelos, Sandra Avila ja Eduardo Valle. Data Augmentation for Skin Lesion Analysis. 2018. arXiv:1809.01442 [cs.CV] https://arxiv.org/abs/1809.01442v1

Xu Zheng, Tejo Chalasani, Koustav Ghosal, Sebastian Lutz ja Aljosa Smolic. STaDA: Style Transfer as Data Augmentation. 2019. 14th International Conference on Computer Vision Theory and Applications, 2019. arXiv:1909.01056 [cs.CV] https://arxiv.org/abs/1909.01056v1

Veit Sandfort, Ke Yan, Perry J. Pickhardt ja Rondald M. Summers. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. 2019. *Sci Rep* 9**,** 16884 (2019). https://doi.org/10.1038/s41598-019-52737-x

Jerry Wei, Arief Suriawinata, Louis Vaickus, Bing Ren, Xiaoying Liu, Jason Wei ja Saeed Hassanpour. Generative Image Translation for Data Augmentation in Colorectal Histopathology Images. 2020. Proceedings of Machine Learning Research 116:10–24, 2020, Machine Learning for Health (ML4H) at NeurIPS 2019

Patrick Helber, Benjamin Bischke, Andreas Dengel ja Damian Borth. Neural Style Transfer for Remote Sensing. 2019. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019

Fabio Henrique Kiyoiti dos Santos Tanaka ja Claus Aranha. Data Augmentation Using GANs. 2019. arXiv:1904.09135v1 [cs.LG]. https://arxiv.org/abs/1904.09135

Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle ja Ole Winther. Autoencoding beyond pixels using a learned similarity metric. 2016. arXiv:1512.09300v2. https://arxiv.org/abs/1512.09300

Joffery L. Leevy, Taghi M. Khoshgoftaar, Richard A. Bauder ja Naeem Seliya. A survey on addressing high-class imbalance in big data. 2018. Journal of Big Data 5**,** 42 (2018). https://doi.org/10.1186

Martin Arjovsky, Soumith Chintala ja Léon Bottou. Wasserstein GAN. 2017. arXiv:1701.07875v3 https://arxiv.org/abs/1701.07875

Patrick Helber, Benjamin Bischke, Andreas Dengel ja Damian Borth. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. 2017. 10.1109/JSTARS.2019.2918242. https://www.researchgate.net/publication/319463676_EuroSAT_A_Novel_Dataset_and_Deep_Learning_Benchmark_for_Land_Use_and_Land_Cover_Classification

Ali Madani, Mehdi Moradi, Alexandros Karargyris ja Tanveer Syeda-Mahmood. Chest x-ray generation and data augmentation for cardiovascular abnormality classification. 2018. Medical Imaging 2018: Image Processing. vol. 10574, p. 105741M. International Society for Optics and Photonics (2018)

Alec Radford, Luke Metz ja Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2015. arXiv:1511.06434v2. https://arxiv.org/abs/1511.06434v2

DCGAN Tutorial — PyTorch Tutorials 1.10.0+cu102 documentation. Haettu 23.10.2021. https://pytorch.org/tutorials/beginner/dcgan_faces_tutorial.html

GitHub - aitorzip/PyTorch-CycleGAN: A clean and readable Pytorch implementation of CycleGAN. 2017. Haettu 23.10.2021. https://github.com/aitorzip/PyTorch-CycleGAN