

Pauli Törmä

**KATSAUS SUPERRESOLUTION
TUOTTAMISEEN GENERATIIVISILLA
KILPAILEVILLA VERKOILLA**

Kandidaattitutkielma
Informaatioteknologian ja viestinnän tiedekunta
Joulukuu 2022

TIIVISTELMÄ

Pauli Törmä: Katsaus superresoluution tuottamiseen generatiivisilla kilpailevilla verkoilla

Kandidaattitutkielma

Tampereen yliopisto

Tietojenkäsittelytieteiden tutkinto-ohjelma

Joulukuu 2022

Superresoluution (SR) tuottaminen on menetelmä, jossa sumeista, epätarkoista tai pienistä kuvista muodostetaan korkealaatuisempi kuva parannetuilla yksityiskohdilla. Superresoluutioita hyödynnetään eri aloilla, muun muassa lääketieteen kuvantamisessa ja kaukokartoituksessa. Kun kuvauslaitteet tai -tilanteet eivät mahdollista korkealaatuisten kuvien ottamista, voidaan kuvia parantaa superresoluutiolla. *Generatiivinen kilpaileva verkko* (GAN) on neuroverkkoarkkitehtuuri, jolla voidaan tuottaa luonnollisen näköisiä SR-kuvia. Tutkielma on katsaus superresoluutioon ja GAN-verkkojen käyttämisestä superresoluution tuottamiseen. Lisäksi tutkitaan, mitä haasteita erityisesti ihmisen kasvoista tehdyissä suurennoksissa kohdataan.

Työ on kirjallisuuskatsaus, joka jakautuu kolmeen osaan. Ensimmäiseksi määritellään SR-kuvan tuottamistehtävä, sen nykyiset käyttökohteet, lyhyt historia ja niiden mittaamismenetelmät. Toiseksi katsotaan, mitä GAN-verkot ovat ja miten ne toimivat superresoluutioita tuottaessa. Neljännessä luvussa käsitellään ongelmia ja niiden ratkaisuja, mitä erityisesti superresoluutioita ihmisen kasvoista tuottaessa kohdataan. Lähteenä käytetään alan artikkeleita tieteellisistä tietokannoista. Artikkeleiksi on koitettu valita melko uusia julkaisuja, koska alalla tapahtuu paljon kehitystä vuosittain.

Työssä on havaittu, että superresoluutio on huonosti asetettu ongelma, mikä tekee sen ratkaisemisesta haastavaa. Koska yhdelle huonolaatuiselle kuvalle voi muodostaa loputtoman määrän erilaisen kaltaisia paranneltuja kuvia, on tärkeä mitata niiden laatua oikeilla mittareilla. Varsinkin ihmisten kasvoista SR-kuvia tuottaessa luonnollisen näköiset tulokset ovat usein kaikkein tärkeimpiä, eivätkä esimerkiksi alhaisen kohinan määrä. GAN-verkot on lähteissä yhtenevästi todettu hyviksi tavoiksi tuottamaan superresoluutioita, johtuen niiden kyvystä tuottamaan koulutusdatan kaltaista tulostetta. GAN-verkkojen superresoluution tuottamisen suurimmiksi ongelmiksi osoittautui käyttökohteita vastaavan koulutusdatan puute ja kuvanlaadun mittaamisen rajoitteet. Ratkaisuiksi ongelmiin esitetään esimerkiksi neuroverkkojen käyttämisestä kattavamman koulutusdatan muodostamiseen.

Avainsanat: Superresoluutio, generatiiviset kilpailevat verkot, kasvojen hallusinaatio, SISR, tekoäly, syväoppiminen

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

SISÄLLYSLUETTELO

1	Johdanto	1
1.1	Tutkimusmenetelmä	1
2	Superresoluutio.....	2
2.1	Superresoluutio yleisesti	2
2.2	Superresoluution kaavallinen esitys	3
2.3	Vanhat menetelmät	5
2.4	Kuvanlaadun mittaaminen	5
2.5	Kasvojen superresoluutio	8
3	Generatiivinen kilpaileva verkko superresoluution tuottamisessa.....	9
3.1	GAN-verkon toiminta superresoluution tuottamisessa	9
3.2	Ongelmia GAN-verkoissa	11
3.3	GAN-verkot kasvojen superresoluutiassa	12
4	Kasvojen superresoluution haasteita ja ratkaisuja	12
4.1	Identiteetin säilyttämisen haaste	13
4.2	Erilaisten kuvakulmien puutteet koulutusdatassa	13
4.3	Käyttökohteesta olevan koulutusdatan puute	14
4.4	Esikäsittelyn hyöty	14
4.5	Laadunmittaamisen kehittäminen	15
5	Keskustelu	16
6	Yhteenveto.....	17
	Lähdeluettelo.....	19

1 Johdanto

Huonolaatuisen kuvan muuttaminen hyvälaatuiseksi kuvaksi selkeämmillä yksityiskohdilla on prosessi, jota kutsutaan superresoluution tuottamiseksi. Superresoluutiolla on nykypäivänä useita käyttökohteita, kuten lääketieteen kuvantamisessa, kuvakartoituksen analysoinnissa ja kasvojentunnistuksessa. Historialliset menetelmät superresoluutioon ovat olleet nopeita ja helppoja, mutta ne tuottavat sumuisia tuloksia vähänkään suuremmissa suurennoksissa. Kaukokartoituksessa kameroiden etäisyyden takia voi olla vaikeaa tulkita pieniä yksityiskohtia kuvista, mutta kuvia ei voida ottaa lähempää. Turvallisuuden kameravalvonnassa, jos haluttaisiin päivittää kaikki kamerat ja laitteet tuottamaan parempaa laatua, vaatisi se isoja rahainvestointeja niin kameroihin kuin tuotetun datan säilyttämiseen ja prosessointiin. Superresoluutio on suunniteltu esimerkiksi näihin käyttötarkoituksiin saamaan halutut selkeät kuvat ilman esitettyjä haasteita.

Superresoluutioiden tuottamiseen on alettu viime vuosina hyödyntämään syväoppimisen menetelmiä, joista generatiivisten kilpailevien verkkojen käyttö on ollut tutkijoiden kiinnostuksessa. Nämä pystyvät tuottamaan luonnollisen näköisiä tuloksia, jotka omaavat enemmän teräviä yksityiskohtia verrattuna aikaisempiin menetelmiin. Tämä on erityisen tärkeää tuotettaessa kasvokuvista SR-kuvia. Verkkojen optimointi tuottamaan parempia tuloksia on vielä tutkijoiden puheenaiheena, joihin on esitetty erilaisia ratkaisuja.

Tutkielman tavoitteena on tehdä katsaus superresoluution, niitä tuottaviin generatiivisiin kilpailullisiin verkkoihin ja kasvojen superresoluutiota muodostaessa tarvittaviin huomioihin. Tutkimuskysymys on, mitä ovat superresoluutio ja GAN-verkot, miksi GAN-verkot soveltuvat superresoluutioon ja mitä haasteita näissä, varsinkin kasvojen suurennoksiin liittyen, on olemassa.

1.1 Tutkimusmenetelmä

Tämä tutkimus on kirjallisuuskatsaus generatiivisiin kilpaileviin verkkoihin ja niiden käyttämisestä superresoluutiokuvien tuottamiseen. Työssä selvitetään miten verkot toimivat, miten superresoluution tuottaminen tehdään ja mitä erilaisia haasteita erityisesti kasvojen superresoluutiossa kohdataan. Työssä hyödynnetään aineistona artikkeleita ja konferenssiartikkeleita alan toimijoilta. Aineisto on haettu pääosin Tampereen Yliopiston

kirjaston pääsilylinkeistä IEEE Xploreen, ScienceDirectiin ja ACM Digital Libraryyn. Hakananoina katsauksessa on käytetty ”generative adversarial network”, ”super-resoluution”, ”face hallucination”, ”SISR” ja ”artificial intelligence”.

Lähteiden valitsemisessa on keskitytty tuoreisiin artikkeleihin. Lähteiksi on koitettu valita monta artikkelia, jotka keskittyvät ihmisten kasvojen superresoluution ongelmaan. Lähteissä on yksimielisesti yhdytty generatiivisten kilpailevien verkostojen hyötyihin superresoluutioiden tuottamiseen, varsinkin ihmisten kasvojen kanssa.

2 Superresoluutio

Tässä kappaleessa esitetään superresoluution määrittely ja sen yleiset käyttötarkoitukset. Superresoluution tuottamisen historiasta esitetään muutamia vanhoja menetelmiä ja niiden rajoitteita. Sitten superresoluution laadun mittaamisen yleisimmät menetelmät ja niiden kaavat esitetään lyhyesti. Viimeiseksi käsitellään kasvojen superresoluutiota, jonka ongelmiin paneudutaan neljännessä luvussa.

2.1 Superresoluutio yleisesti

Superresoluution (super-resolution, SR) tuottaminen on tietokonenäön ja kuvaprosessin menetelmä, joka parantaa *huonolaatuisten* (low resolution, LR) kuvien visuaalista havaitsemista, tuottaen niistä *hyvälaatuisia* (high resolution, HR) kuvia (Lepcha et al., 2022). Superresoluution tarkoituksena on muuttaa sumuiset, epäselvät ja pieniresoluutioiset kuvat selviksi hyvälaatuisiksi kuviksi parannetuilla yksityiskohdilla. Reaali maailmassa kuvien laatua laskee esimerkiksi kamerasensorien kohina, kuvien häviölliset pakkausmenetelmät, hitaasta suljinnopeudesta johtuva liikesumennus ja huono kameran linssin tarkennus. Yhtä tai useampaa kuvaa voidaan hyödyntää suurennoksen tekoon, mutta tässä tekstissä keskitytään yhtä kuvaa käyttäviin menetelmiin. Tätä kutsutaan *yhden kuvan superresoluutioksi* (single image super-resolution, SISR). Viime vuosina superresoluution menetelmät ovat saaneet tieteellisten tutkijoiden huomion ja varsinkin lääketieteen käyttötarkoituksiin löytyy paljon papereita tieteellisistä tietokannoista.

Superresoluutiolla parannetuista kuvista saa selkeämpää informaatiota sijainneista ja luontaisista esineistä, mistä on hyötyä useissa käyttötarkoituksissa. Superresoluutiolla on tärkeä rooli monella alalla, kuten lääketieteellisessä diagnostiikassa, kaukokartoituskuvauksen analysoinnissa ja kasvojentunnistuksessa (Lepcha et al., 2022). Varsinkin lääke-

tieteen kuvantamisessa ja turvallisuuden valvonnassa tarvitaan superresoluutiomenetelmiä, johtuen näiden käyttökohteiden tuottamien kuvien rajoitteista. Paremman kuvan saamiseksi ilman superresoluutiota täytyisi investoida suuria summia parempiin kuvauslaitteisiin. Kameravalvonnassa tarvittavien kameroiden määrän ja niiden tuottaman datan prosessoinnin takia se ei aina ole käytännöllistä, eikä lääketieteessä parempia kuvantamislaitteita välttämättä ole olemassa. Superresoluution hyödyntäminen vanhojen kuvien entisöintiin tai suoratoistettavan videon parantaminen kaistanleveyden säästämiseksi ovat esimerkkejä arkisemmista käyttökohteista.

Monet ihmiset voivat arjessaan jo hyötyä superresoluutiomenetelmistä esimerkiksi videopelejä pelatessa. Tietokoneiden näyttöjen resoluutioiden kasvaessa ja videopelien grafiikkokojen parantuessa ovat laskennalliset tarpeet niiden suorittamiseen nousseet valtavasti. Esimerkiksi päivittäessä 1080p-näytön 4k-näytöksi nousee pikselien määrä nelinkertaiseksi, mikä tarkoittaa nelinkertaista prosessoinnin tarvetta. Grafiikkaprosessoreiden on vaikea kehittyä näin paljon lyhyessä ajassa, joten kaikki merkittävät grafiikkaprosessoreiden valmistajat ovat viime vuosina kovaa tahtia kehittäneet tuotteisiinsa superresoluutioiden menetelmiä. Niitä käyttämällä voidaan peli suorittaa pienemmällä resoluutiolla, josta muodostetaan reaaliajassa superresoluutiota näytölle. Nykyään siis jokainen kohdullisen uuden näytönohjaimen omaava ihminen voi kokea superresoluution hyötyjä pelejä pelatessaan. Yhtiöt eivät kuitenkaan avoimesti kerro käyttämiään menetelmiä, joten näistä ei ole tutkimuspapereita saatavilla. Yksi esimerkki näistä on Nvidian *DLSS* (deep learning super sampling), joka yhtiön mukaan tuottaa tekoälyn avulla superresoluutiota (Nvidia, 2022).

2.2 Superresoluution kaavallinen esitys

Superresoluution tuottaminen on klassinen *huonosti asetettu inversio-ongelma* (ill-posed problem). Tämä tarkoittaa, ettei siihen ole aina ratkaisua, ratkaisu ei ole yksiselitteinen tai ratkaisu ei riipu jatkuvasti annetusta datasta. Yhdelle LR-kuvalle voi olla olemassa rajattoman monta eri HR-kuvaa, riippuen pienistä variaatioista värin, kirkkauden, kuvakulman tai muitten muuttujien takia. Tämän takia superresoluution suorittaminen on vieläkin haaste tietokonenäössä, vaikka sitä on tutkittu vuosikymmeniä.

Ongelma voidaan esittää Kaavan (1) mallilla, missä huonontamisfunktio d :n tuloste I_x on LR-kuva. I_y on alkuperäinen syötteenä ollut HR-kuva (viitekuva) ja δ on syöteparametrit huonontamisfunktiolle, kuten skaalausta, sumentamisen tyyppiä ja kohinaa. Tämä huonontamisprosessi on tuntematon suurennosta tehdessä ja vain LR-kuvia käytetään muodostamaan HR-kuvia superresoluutiassa.

$$(1) I_x = d(I_y; \delta)$$

Superresoluution prosessissa yritetään saada alkuperäisestä HR-kuvasta I_y sitä arviolta vastaava SR-kuva \hat{I}_y Kaavan (2) mallilla, missä F on superresoluution malli ja θ kuvastaa sen parametrejä. Tässä siis pyritään kumoamaan huonontamisfunktion vaikutukset kuvalle. Koska superresoluution ongelma on huonosti asetettu ongelma, on HR-kuvalla (I_y) loputon määrä Kaavan täyttäviä \hat{I}_y mahdollisuuksia (Wang et al. 2021).

$$(2) \hat{I}_y = F(I_x; \theta)$$

Superresoluution tarkoituksena on minimoida häviöfunktio, kuten Kaavassa (3) on esitetty. $\mathcal{L}(\hat{I}_y, I_y)$ on häviöfunktio tuotetun SR-kuvan ja oikean HR-kuvan välillä, h on kompromissiparametri ja $\psi(\phi)$ on regularisointitermi. Kaikista yleisin häviöfunktio superresoluutiassa on pikselipohjainen *keskimääräinen neliövirhe* (mean square error, MSE), jota kutsutaan myös pikselihäviöksi, mutta parhaiten toimivat mallit käyttävät useiden häviöfunktioiden yhdistelmiä (Wang et al. 2021). Näiden eri häviöfunktioiden yhdistely ja optimointi on superresoluution alalla tutkijoiden papereissa suuressa roolissa.

$$(3) \hat{\phi} = [\min \mathcal{L}(\hat{I}_y, I_y)]_{\phi} + h\psi(\phi)$$

Häviöfunktioita superresoluution alalla käytetään arvioimaan uudelleenrakennusvirheitä ja ohjaamaan mallin optimoimista. Erilaisia häviöfunktioita voidaan käyttää tuottamaan realistisempia ja laadukkaampia tuloksia. Näitä ovat esimerkiksi tekstuuri-, kilpailu- ja sisältöhäviö. Alun perin tutkijat usein käyttivät pikselikohtaista häviötä mittaamiseen, mutta myöhemmin huomasivat, ettei se mittaa uudelleenrakentamien laatua tarkasti (Wang et al., 2020). Häviöfunktioit mittaavat tuotetun superresoluution onnistumista säilyttämään esimerkiksi samanlaiset värit ja tekstuurit alkuperäiseen verrattuna. Muuttamalla häviöfunktioita saadaan malli tuottamaan halutunlaista tulostetta.

2.3 Vanhat menetelmät

Ennen koneoppimisen menetelmiä superresoluutioiden tuottamiseen, suurennoksien tuottamiseen on käytetty interpolointimenetelmiä. Kuvien interpolointi, eli kuvien skaalaus, tarkoittaa digitaalisten kuvien koon muuttamista, mitä hyödynnetään yleisesti kuviin liittyvissä ohjelmissa (Wang et al., 2020). Klassiset menetelmät, jotka perustuvat interpolointiin, ovat muun muassa lähinnaapuri-interpolointi, bilineaarinen interpolointi ja bikuutiointerpolointi. Interpolointi parantaa kuvan resoluutiota perustuen sen omiin kuvasignaaleihin tuomatta siihen yhtään uutta informaatiota (Cao et al., 2021). Tämän takia niiden tulokset omaavat sivuvaikutuksia, kuten kohinan lisääntymistä ja yksityiskohtien sumentumista. Interpoloinnin keinot ovat olleet helppoja ja luotettavia reaaliaikaiseen superresoluutioiden tuottamiseen jo vuosikymmeniä, mutta nykyään oppimispohjaiset algoritmit menestyvät paremmin. Tämän takia nykyinen trendi on korvata interpolointiin pohjautuvia suurennosmetodeja koneoppimiseen pohjautuvilla.

Viime vuosikymmenellä yhden kuvan superresoluutioiden tuottaminen *konvoluutioverkoilla* (convolutional neural network, CNN) on tuottanut vakuuttavia tuloksia (Singla et al., 2022). Ensimmäinen esitetty syväoppimisen superresoluutiomenetelmä oli *SRCNN* (super-resolution convolutional neural network). CNN-verkot keskittyvät minimoimaan häviön todellisen suuriresoluutioisen kuvan ja tuotetun suuriresoluutioisen kuvan välillä, välittämättä kuvien havainnollisesta laadusta. Ne pystyvät saavuttamaan hyvät SSIM- ja PSNR-arvot, jotka määritellään seuraavassa aliluvussa, mutta havainnollinen laatu voi silti jäädä alhaiseksi. Parantaakseen CNN-verkkojen havainnollista laatua, tutkijoiden mielenkiinnon herättäneet generatiiviset kilpailulliset verkot on esitetty ratkaisuksi. Nämä voivat tuottaa ihmiselle miellyttäviä kuvasuurennoksia, jotka silti pitävät häviön pienenä (Singla et al., 2022). Kolmannessa luvussa kerrotaan GAN-verkkojen toiminnasta.

Viimeisimmät algoritmit superresoluutioiden tuottamiseen ovat suoriutuneet paremmin kuin vanhat menetelmät, mutta niillä on myös useita rajoitteita ja haasteita. Nämä ovat kuitenkin yleisessä ymmärryksessä ja ohjaavat tulevaisuuden tutkimusten suuntaa (Lepcha et al., 2022). Neljännessä luvussa esitetään lähteissä nousseita ongelmia ja haasteita, joita koetaan erityisesti ihmisten kasvoista tuotettaviin SR-kuviin.

2.4 Kuvanlaadun mittaaminen

Kuvanlaatu on termi, joka tarkoittaa kuvan visuaalisten merkittävien ominaisuuksien yhdistelmää. *Kuvanlaadun arviointi* (Image quality assessment, IQA) on tekniikka, joka

yleisesti yrittää sisällyttää kvalitatiivisia tekniikoita kuvien arviointiin. Laadunmittaamiseen voidaan joko käyttää tai olla käyttämättä viitekuvia. IQA-metodit voidaan jakaa subjektiivisiin metodeihin (ihmisen käsitys kuvasta, eli onko kuva luonnollinen ja hyvälaatuinen) ja objektiivisiin metodeihin (kvalitatiiviset metodit, missä kuvanlaatu voidaan koneellisesti laskea numeroksi) (Wang et al. 2021). Erilaiset objektiivisiksi kehitetyt mitausmenetelmät eivät kuitenkaan välttämättä kuvasta miten todellisuudessa ihmiset kokevat SR-kuvan laadun. Tapoja kuitenkin käytetään, koska on laskennallisesti kalliimpaa mitata, kuinka realistiselta kuva näyttää ihmiselle (Lepcha et al. 2022). Objektiivisiä mitaustapoja eri mallien suoriutumiseen tarvitaan niiden kouluttamisessa, monitoroinnissa ja muihin malleihin vertaamisessa.

Superresoluutioiden laatua mitataan yleensä *juuren keskimääräisellä neliövirheellä* (root mean square error, RMSE), *huippusignaali-kohinasuhteella* (peak signal to noise ratio, PSNR) ja *rakenteen samankaltaisuusindeksillä* (structure similarity index measure, SSIM). Nämä ovat likiarvoja käsitykseen suurennoksen laadusta.

PSNR on mittaustekniikka, jossa analysoidaan signaalin voimaa verrattuna kohinan voimaan. Varsinkin kuvissa tätä käytetään mittaamaan kvalitatiivisesti kuvanpakkauksen laatua. Superresoluutiossa kuvan PSNR määritetään maksimi pikseliarvolla ja keskimääräisellä neliövirheellä viitekuvan ja SR-kuvan välillä, jota kutsutaan myös kuvanvääritymäkohinan voimaksi.

PSNR määritellään mustavalkoiselle kuvalle Kaavalla (4) ja MSE Kaavalla (5). Maksimi pikseliarvo on M , viitekuva on I_r , t on pikselien määrä ja SR-kuva on I_y . i ja j ovat pikselien x-y koordinaatteja. M on kuvan maksimaalinen mahdollinen pikseliarvo, eli kahdeksanbittisessä kuvassa sen maksimiarvo on 255. MSE-kaava vertaa kahden kuvan pikselien värieroja. Värillisissä kuvissa MSE lasketaan käyttämällä jokaisen pikselin RGB-arvoja. PSNR:n Kaavasta näkee, että se on relaatiossa SR-kuvan ja viitekuvan yksittäisiin pikselien intensiteettiarvoihin. PSNR-arvo voi joissain tilanteissa olla harhaanjohtava, koska SR-kuva ei välttämättä näytä samanlaiselta viitekuvan kanssa. Superresoluution alalla PSNR-arvoa käytetään vielä yleisesti vertamaan uusia SR-menetelmiä aiemmin julkaistuihin.

$$(4) \text{ PSNR} = 10 \log_{10} \left(\frac{M^2}{\text{MSE}} \right)$$

$$(5) \text{ MSE} = \frac{1}{t} \sum_{i=1}^t (I_r(i) - I_y(i))^2$$

Koska ihmisten näkö on tehokas erottamaan rakenteellisen tiedon kuvasta, eikä PSNR arvossaan välitä kuvan rakenteellisesta kompositiosta, SSIM esitettiin mittaamaan tätä rakenteellista yhteneväisyyttä kuvien välillä (Wang et al. 2021). Tämä tapahtuu vertaamalla kontrastia, luminanssia ja rakenteellisia yksityiskohtia viitekuvaan. Superresoluutiossa tuotettua SR-kuvaa verrataan koulutusdatan alkuperäiseen kuvaan. Kaikenlaisiin tilanteisiin SSIM ei kuitenkaan toimi, koska esim. lääketieteen kuvantamisessa varianssi ja luminanssi viitekuvaan ovat pieniä, joten SSIM voi olla epävakaa, joka johtaa virheellisiin tuloksiin. Kuitenkaan luonnollisiin kuviin tämä ei päde. Koska SSIM mittaa rekonstruktion laatua ihmisen näköä matkien objektiivisesti, se paremmin saavuttaa havaitun laadun mittaamisen tarpeet, mistä seuraa sen laaja käyttö (Wang et al., 2021).

SSIM voidaan esittää Kaavan (6) tavalla, missä SSIM koostuu kahden näytteen välisistä kolmesta mittauksesta: Com_l luminanssi, Com_c kontrasti ja Com_s koostumus. α , β ja γ kuvaavat painotuksia eri mittareille, joita muuttamalla voidaan määrittää tietyn arvon tärkeys. I_r on alkuperäinen HR-kuva ja \hat{I} on tuotettu SR-kuva. Kaavasta saadaan SSIM-arvo, joka on desimaaliarvo välillä -1–1. Yksi kuvastaa täydellistä samankaltaisuutta, nolla samankaltaisuuden puuttumista ja miinus yksi täydellistä antikorrelaatiota.

$$(6) \quad SSIM(I_r, \hat{I}) = \{Com_l(I_r, \hat{I})\}^\alpha \{Com_c(I_r, \hat{I})\}^\beta \{Com_s(I_r, \hat{I})\}^\gamma$$

Yleisesti käytetty subjektiivinen kuvanlaadunmittaustapa on *keskimääräinen mielipidepisteitys* (mean opinion score, MOS) (Wang et al. 2021). Siinä ihmisarvostelijat arvioivat kuvat asteikolla 1–5 ja lopullinen MOS-arvo lasketaan kaikkien arvosteluiden keskiarvoksi. Vaikka MOS kärsii subjektiivisten arvosteluiden ongelmista, kuten arvostelijoiden kriteerien varianssista, pystytään sillä mittaamaan hyvin ihmisten havainnollista laatua suurennoksista. Saadut tulokset on kuitenkin vaikea tuottaa uudestaan ja verrata objektiivisiin kuvanlaadun mittareihin.

Ihmisen arviointia matkivia syväoppimisen malleja on esitetty, ja ne pystyvät tuottamaan lupaavia tuloksia, mutta ne vaativat toimiakseen paljon resursseja, kuten suuria tietojoukkoja koulutukseen. Myös minkälaista laatua suurennoksilta halutaan, on avoin kysymys, eli halutaanko enemmän luonnolliselta näyttäviä kuvia vai enemmän viitekuvan kaltaisia kuvia. Koska tämä kysymys on vielä tutkijoilta yleisesti vastaamatta, PSNR ja SSIM ovat pääosin vakiintuneet kuvanlaadun mittaamiseen superresoluutiossa. Kuitenkin tulevaisuudessa on tarvetta kehittää parempia laadunmittauskeinoja, jotta voidaan objektiivisemmin mitata eri malleja keskenään.

2.5 Kasvojen superresoluutio

Kasvojen superresoluutio on esiin nousemassa oleva ala, jossa tähdätään tehdä hyvälaatuisia kasvokuvia pieniresoluutioisista kuvista. Superresoluutiotekniikoita, jotka erikoistuvat kasvoihin, kutsutaan myös yleisesti *kasvojen hallusinaatioksi* (face hallucination). Kasvokuvien prosessoinnissa käytettävät SR-verkot ovat yleensä muita monimutkaisempia, koska kasvojen uudelleenrakennuksessa täytyy koittaa säilyttää kasvojen identiteetti. Vaikka tuotettu SR-kuva ei olisi henkilön tunnistamisen kannalta kriittinen, ovat ihmiset hyviä huomaamaan kasvoissa olevia epämääräisyyksiä helpommin, kuin muissa kuvien osissa. Pienet geometriset vääristymiset suussa ja silmissä voivat vain vähäisesti laskea suurennoksen objektiivista laatua, mutta voivat merkittävästi laskea subjektiivista koettua laatua (Hsu et al., 2019). Näin ollen yleinen kasvojen muoto ja tekstuuri, sekä lokaalit geometriset muodot (suu, nenä, silmät) täytyy kasvojen superresoluutiossa harkita tarkkaan. Kasvojen osista ihmisen tunnistamisesta tärkeimpiä ovat silmät, nenä ja suu (Zhang et al., 2021). Yleisesti kasvojen SR-tekniikat sisältävät aiempaa tietoa kasvoista auttaamaan tuottaa oikein muodostettuja kasvokuvia. Tällaista tietoa käyttävät menetelmät usein lisäävät prioriverkon pääverkkoon, mikä lisää parametrien määrää, joka taas lisää tietojenkäsittelyn resursseja (Zhao et al., 2022).

Kasvokuvat sisältävät henkilön kasvoista uniikin biologisen tiedon, joista voidaan tunnistaa yksilö. Monilla aloilla hyödynnetään kasvojentunnistusta, kuten turvallisuudessa, lääketieteessä ja julkisessa liikenteessä. Kuitenkin näissä käytettävät kuvanottojärjestelmät usein tuottavat huonolaatuisia kuvia laitteiston rajoitteiden, kohteiden liikkeen ja etäisyyden takia. Kuvat ovat sumuisia, omaavat sensorikohinaa ja kasvot pieniä, johtuen mm. kameroiden etäisyyksistä. Tämä johtaa haasteisiin niiden hyödyntämisestä kasvojentunnistuksessa (Cao et al., 2021). Empiiriset tutkimukset kasvojentunnistuksen alalla ovat osoittaneet, että pienin resoluutio kasvokuvissa tehokkaaseen tunnistukseen on oltava vähintään välillä 32 x 32 ja 64 x 64 pikseliä (Hsu et al., 2019).

Useat tutkimukset ovat esittäneet erilaisia metodeja parantamaan kasvojen superresoluutiota. Nämä yleensä tekevät verkoista monimutkaisia ja täten vaativat paljon muistia ja tietojenkäsittelyn resursseja (Zhao et al., 2022). Lisääntyneet tietojenkäsittelyn tarpeet voivat tehdä niistä vaikeita implementoida yleisiin käyttötarkoituksiin ja varsinkin reaaliaikaisiin kohteisiin. Luvussa neljä paneudutaan erityisesti kasvojen superresoluutioita tuottaessa kohdattaviin ongelmiin.

3 Generatiivinen kilpaileva verkko superresoluution tuottamisessa

Tässä luvussa esitellään GAN-verkot ja niiden toiminta superresoluutioiden tuottamiseen. Sen jälkeen lyhyesti GAN-verkkojen yleisiä ongelmia. Viimeiseksi verkkojen käytöstä kasvojen superresoluutioon käydään läpi.

Generatiivinen kilpaileva verkko (generative adversarial network, GAN) on viime vuosina tutkijoiden kiinnostuksen herättänyt neuroverkkoarkkitehtuuri, jota ensimmäisen kerran ehdotti Ian J. Goodfellow vuonna 2014. Nykyään kuva- ja näköala saa GAN-tutkijoilta eniten huomiota, johtuen verkkojen kyvystä tuottaa luonnollisemman tyyllisiä tuloksia, kuten fotorealistisia kuvia. Alun perin GAN-verkot esitettiin ohjaamattoman oppimisen muodoksi, mutta ne ovat osoittautuneet sopiviksi myös valvottuun tai puolivalvottuun oppimiseen. Superresoluution tuottamisen lisäksi käyttökohteita GAN-verkoille ovat muun muassa kuvien tuottaminen tekstisyötteen pohjalta ja semanttinen kuvien editointi.

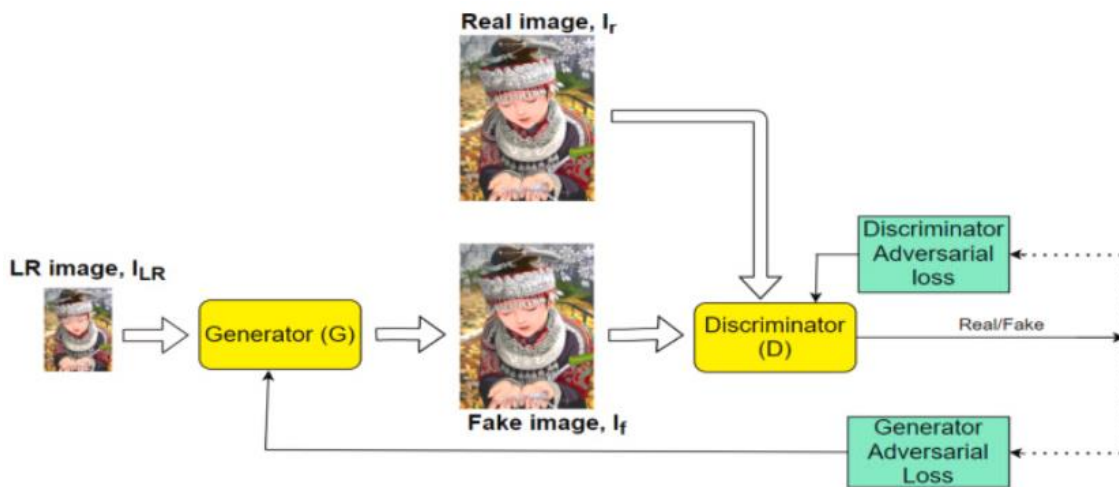
3.1 GAN-verkon toiminta superresoluution tuottamisessa

Kilpailevan nimikkeen GAN-verkko saa, koska niissä on kaksi eri neuroverkkoa kilpailuasetelmassa. Vertauskuvana GAN-verkon toiminnalle superresoluution tuottamisessa on mahdollista ajatella toista verkkoa taiteen väärentäjänä ja toista taide-eksperttinä (Creswell, et al. 2018). Väärentäjä on nimeltään generaattori, joka tuottaa taideväärennöksiä, tavoitteenaan tehdä realistisia kuvia. Ekspertti on nimeltään diskriminaattori, joka saa syötteekseen oikeita kuvia ja generaattorin väärennöksiä, tavoitteenaan erottaa ne toisistaan. Molempia verkkoja koulutetaan vuorotellen, jotta saadaan yhä parempia tuloksia. Generatiiviset mallit oppivat koulutusdatan statistisen jakauman, joista ne tuottavat uudenlaista dataa (Creswell, et al. 2018). Tämän takia koulutusdatan valinnalla on tärkeä rooli verkon kykyyn tuottaa hyvää dataa ja sen kykyä menestyä reaali maailman käyttökohteissa.

Generaattorilla ei ole pääsyä oikeisiin kuviin, joten se oppii vain vuorovaikutuksestaan diskriminaattorin kanssa. GAN-verkkojen kouluttaminen pitää sisällensä parametrien etsimisen diskriminaattorille, jotka maksivoivat sen tarkkuuden löytää generoituja kuvia, ja parametrien etsiminen generaattorille, jotta se maksimimaalisesti huijaa diskriminaattoria.

Generaattori G oppii ja arvioi koulutusdatajakauman, tuottaakseen uusia näytteitä arvioidusta jakaumasta (Singla et al., 2022). Diskriminaattori D ottaa tuotetun kuvan syötteenä ja arvioi mahdollisuuden (P_D), että tuotettu kuvat ($p(data)$) kuuluvat kohdekuvajakaumaan ($p(model)$), joka esitetään Kaavassa (7). Kuvassa 1 esitetään lohkokaavio GAN-verkosta, jossa verkon generaattori koulutetaan tuottamaan SR-kuvia oppimalla koulutusdatasta kilpailullisella häviöllä.

$$(7) P_D = \frac{p(data)}{p(data) + p(model)}$$



Kuva 1- GAN-verkon lohkokaavio. (Singla et al., 2022).

Generaattori on optimaalisin, kun diskriminaattori ei erota oikeita kuvia generoiduista, eli generaattori tuottaa näytteen, jolle diskriminaattori antaa mahdollisuudeksi (P_D) 0,5, eli diskriminaattori on 50 % varma sen aitoudesta/epäaitoudesta. Tällöin generaattori on tuottanut näytteen, joka kuuluu opetusdatan jakaumaan. Tätä tasapainotilaa kutsutaan Nashin tasapainoksi, joka on peliteoria, jossa kumpikaan osapuoli ei voi yksipuolisesti muuttaa strategiaansa saavuttaakseen paremman lopputuloksen itselleen. (Singla et al., 2022). Verkon kouluttamisen jälkeen diskriminaattori poistetaan ja sillä voidaan tuottaa haluttua tulostetta.

Tekniikka kouluttaa GAN-verkkoja tuottamaan uusia näytteitä on kilpailullista kouluttamista. Se koostuu kahdesta vaiheesta: Generaattorin parametrit pidetään vakiona ja diskriminaattoria koulutetaan erottamaan oikeat ja tuotetut jakaumat. Diskriminaattorin parametrit pidetään vakiona ja generaattoria koulutetaan kuvilla, jotka kuuluvat kohde jakaumaan (Singla et.al, 2022). Koulutuksen tavoitteena on estimoida oikean datan jakauma ja tuottaa näytteitä, jotka kuuluvat siihen jakaumaan. GAN-verkon kouluttamisessa on löydettävä diskriminaattorin parametrit, jotka maksivoivat sen kyvyn luokitaa

väärä ja oikea data, sekä generaattorin parametrit, jotka maksimaalisesti huijaavat diskriminaattoria (Creswell, et al. 2018).

Standardi GAN-verkon häviöfunktio, joka tunnetaan myös nimellä min-max häviö, voidaan esittää kaavalla (8), missä x saadaan näytteenotolla oikean datan jakaumasta tai $p(x)$, z saadaan näytteenotolla aiemmasta datasta tai $p(z)$, $E[x]$ on odotettu arvo mille tahansa satunnaiselle muuttujalle x , $D(x)$ on mahdollisuus, että x on otettu oikeasta datasta eikä generoidusta datasta. Diskriminaattorin päätavoite on maksimoida tavoitefunktio minimoimalla $D(G(z))$ ja maksimoimalla $D(x)$, joka merkitsee, että diskriminaattori luokittelee oikeaa dataa oikeaksi ja tuotettua dataa tuotetuksi. Generaattorin päätavoite on minimoida tavoitefunktio maksimoimalla $D(G(z))$, joka tarkoittaa, että diskriminaattori luokittelee tuotettua dataa oikeaksi (Kumar, Dhawan, 2020).

$$(8) \min_G \max_D V(D, G) = E_{x \sim p(x)}[\log D(x)] + E_{z \sim p(z)}[\log(1 - D(G(z)))]$$

3.2 Ongelmia GAN-verkoissa

Lupaavien tulostensa lisäksi GAN-verkkojen koulutuksessa kohdataan myös useita haastavia ongelmia, joita on tässä esitetty lyhyesti muutama. *Mallin romahduksessa* (Mode collapse) verkko jumittuu tiettyyn paikalliseen minimiin, jossa se pystyy kaappaamaan vain pienen osan jakaumasta. Tämä tapahtuu, kun generaattori tuottaa samankaltaista tulostetta pienellä varianssilla. Usein halutaan, että verkko tuottaa paljon erilaisia mahdollisia tuloksia, eikä vain yhtä samaa. Generaattori ei siis uskollisesti mallinna kohdejakautaa ja ei käytä osaa siitä. *Katoava gradientti* (vanishing gradient) estää generaattoria oppimasta mitään, koska diskriminaattorista on tullut liian tehokas. Kaikki tuotetut kuvat luokitellaan tekaistuiksi, eikä generaattori pysty parantamaan tulostettaan. *Tasapainon epäonnistumisessa* (convergence failure) GAN-verkko voi epäonnistua pääsemään Nashin tasapainoon. Tämä voi yleisesti tapahtua, kun generaattori tuottaa huonoja tuloksia, jotka diskriminaattori helposti tunnistaa.

Vaikka GAN-pohjaiset superresoluution menetelmät kärsivät näitä ongelmia, ne ovat huomattavasti parantaneet tuotettujen SR-kuvien visuaalista laatua vanhoihin PSNR-keskeisiin metodeihin (Singla et al., 2022). Verkkoja kouluttaessa täytyy ottaa mahdolliset ongelmat huomioon ja keskeyttää koulutus mahdollisten ongelmien esiintyessä. Myös GAN-verkkojen muistin tarvetta ja energian kulutusta on parannettava, jotta niitä voidaan hyödyntää reaaliaikaisissa käyttökohteissa.

3.3 GAN-verkot kasvojen superresoluutiossa

Useimmat uudet kasvojen superresoluutioiden menetelmät käyttävät GAN-verkkoja, koska aikaisemmat menetelmät keskittyvät pääosin maksimoimaan PSNR-arvoa, mikä johtaa melko sumeisiin kuviin ilman tarkkoja yksityiskohtia. GAN-verkot puolestaan pystyvät tuottamaan luonnollisen näköisiä kuvia parannetuilla yksityiskohdilla. Kuitenkaan GAN-verkkoja ensimmäiseksi esittäneen Goodfellowin alkuperäinen verkko ei olisi kelvannut kasvojen superresoluutioon. Siinä ollut generaattori suunniteltiin ottamaan kohinaa syötteenä ja antamaan tulostetta realistisilla ominaisuuksilla. Tämän satunnaisuus ja sen rajoitteet yhteneväisten kuvaominaisuuksien säilyttämisessä tarkoittaa, että verkko ei pystynyt täyttämään superresoluution tavoitteita (Hall et al. 2020). Esimerkiksi vanhaa elokuvaa entisöidessä olisi superresoluution käyttö turhaa, jos elokuvatähden kasvot näyttäisivät selvemmilta ja luonnollisilta, mutta olisivat aivan eri henkilön näköiset. GAN-verkot ovat kuitenkin muokattavissa superresoluutioiden tuottamiseen. Yksi suosituimmista ja menestyneimmistä GAN-verkoista on *Superresoluutio GAN* (SRGAN). Jotta kasvojen SR-kuvissa saavutettaisiin parempia tuloksia, tällaisia valmiita verkkoja koitetaan jatkokehittää huomiomaan ihmisten kasvot tarkemmin. Kuitenkin suurin osa olemassa olevista kasvojen superresoluutiometodeista keskittyvät vain tuottamaan visuaalisesti miellyttäviä HR-kuvia harkitsematta kasvojen identiteettiä (Hsu et al., 2019).

4 Kasvojen superresoluution haasteita ja ratkaisuja

Tässä luvussa käydään läpi erilaisia haasteita, joita kohdataan superresoluutioiden tuottamisessa GAN-verkoilla. Tarkoitus on keskittyä asioihin, jotka ovat erityisen tärkeitä ihmisten kasvoista tuotetuissa SR-kuvissa. Vaikka monet lähteet käsittelevät superresoluution käyttöä kasvojentunnistamista varten, on tämä tavoite kyseenalainen. GAN-verkko tuottaa puuttuvat yksityiskohdat koulutusdatasta opitulla tavalla eikä sen tuottama kasvo kuvasta todellisuutta. Kuitenkin näitä lähteitä käytetään tutkimuksessa, koska ne koittavat ratkoa ongelmia, joita kohdataan myös muissa superresoluution käyttökohteissa. Jos esimerkiksi entisöidään vanhaa filmivalokuvaa superresoluutiolla, on tärkeää koittaa säilyttää kuvan henkilön identiteetti. Aivan erinäköisten kasvojen tuottaminen superresoluutiolla on välttävää.

4.1 Identiteetin säilyttämisen haaste

Monet kasvojen hallusinointi keinot tuottavat visuaalisesti miellyttäviä kasvokuvia, mutta niiden tarkkuus kasvojentunnistuksessa on alhaisempi kuin odotettu, koska kasvojen rakenne ja niiden yksityiskohtia ei ole hyvin säilytetty (Hsu et al., 2019). Kuvassa 2 on esitetty ero kahden superresoluutiomenetelmän välillä, jossa a on syötekuva 8x8 pikseliä, b on identiteettitietämätön SR-menetelmä, c on identiteettisäilyttävä SR-menetelmä ja d on alkuperäinen kuva. Identiteetistä piittaamattomat mallit voivat tuottaa visuaalisesti parempia kuvia, mutta niitä ei voida hyödyntää kasvojentunnistukseen.



Kuva 2 – Erot menetelmien tulosteessa identiteetin säilyttämisen kannalta. (Hsu et al., 2019)

Jotta voidaan tuottaa identiteetin säilyttäviä suurennoksia, tarvitaan nimetty harjoitusdata oppiminaan identiteettitietoisia esityksiä. Yksi tapa on upottaa täyden luokan tunnusmerkit harjoituskasvoihin, esimerkiksi yksittäisten harjoituskasvojen oikeat identiteetti-luokat. Nimetyt koulutusdatan hankkiminen on haaste, etenkin jos käyttötarkoitus vaatii suuren määrän koulutusdataa useilla eri identiteeteillä. Oikeiden henkilöiden identiteettitietojen lisääminen nostattaa myös yksityisyyskysymyksiä.

4.2 Erilaisten kuvakulmien puutteet koulutusdatassa

Valtavirran kasvojen hallusinaatiomenetelmät keskittyvät tuottamaan suurennoksia kasvokuvista, joissa henkilön kasvot ovat lähes täydellisen suorassa edestä (Zhang et al., 2021). Kuitenkin oikean maailman turvallisuuskameroiden asettelu johtaa kuviin, joissa henkilön kasvot ovat sivusta tai ylhäältä otettuja. Tällaiset kuvat aiheuttavat ongelmia konehavaintoalgoritmeille ja myös ihmisten kyvyille tunnistaa henkilöitä. Kameroita ei voida kuitenkaan monissa tilanteissa siirtää ihmisten kasvojen tasolle, joten ongelma täytyy ratkaista eri tavalla. Tähän ongelmaan on olemassa tekniikkoja, joissa kasvokuvat muunnetaan edestä oleviksi, joita kutsutaan *etukuvakulman syntetisoinniksi* (frontal view synthesis tai face frontalisation). Myös tämä ongelma on huonosti esitetty ongelma, kuten SR-ongelma, joten virheitä kummassakaan prosessissa ei voida eliminoida ja yhdistettynä ne vahvistavat virheitään (Zhang et al. 2021). Metodeja kasvokuvan muuttamisesta edestä otetuksi on jo olemassa, mutta ne vaativat HR-kuvia toimiakseen ja jotta voitaisiin tehdä

kasvojen SR, tarvitaan kasvoista edestä otettu kuva. Naiivi idea on käyttää näitä olemassa olevia menetelmiä peräkkäin, mutta ne tuottavat suuren määrän grafiikkahäiriöitä ja vääristymiä (Zhang et al. 2021). Ongelmaan on esitetty ratkaisuksi muun muassa VividGANia, joka yrittää tehdä kuvakulman muunnoksen ja superresoluution yhdessä asteittain ja koittaen vähentää grafiikkahäiriöitä ja sumuisuutta.

4.3 Käyttökohteesta olevan koulutusdatan puute

Jos verkon koulutuksessa ei ole koskaan käytetty LR-kuvia, jotka näyttävät oikean maailman huonoilta kuvilta, ei verkko pysty tuottamaan hyvää tulostetta harjoitusdatan ulkopuolella. On kuitenkin yleisesti vaikea hankkia suuria määriä paritettuja LR- HR- kuvia oikeista kohteista koneoppimista varten. Tämän takia useimmat kasvojen SR-metodit käyttävät ennalta määritettyjä kuvanlaadun huononnutusprosesseja saadakseen tarvittavat LR-kuvat koulutukseen. HR-kuva esimerkiksi pienennetään ja siihen lisätään kohinaa ja sumuisuutta, josta saadaan HR-kuvalle tarvittava LR-kuva koulutukseen. Harjoitusdata, joka on saatu ennalta määritellyllä huononnutusprosessilla, aiheuttaa negatiivisen vaikutuksen mallin kykyihin, kuten realismiin ja yksityiskohtien puutteen oikeissa käyttökohteissa (Zhao et al., 2022). Tämän takia halutaan käyttää suurta määrää HR-kuvia ja niitä vastaavia LR-kuvia, jotka kärsivät luonnollisesta kuvien laadun huonontumisesta, verkkojen kouluttamiseen. Kuitenkaan luonnollisesta huononemisesta kärsiviä kuvia ja vastaavia HR-kuvia ei ole saatavilla tarvittavia määriä verkkojen koulutukseen. Tämän takia koulutusdatan LR-kuvien hankkiminen GAN-verkolla, joka tuottaa HR-kuvista luonnollisia LR-kuvia, voi auttaa. Verkko koulutetaan syöttämällä HR-kuvia kasvoista ja LR-kuvia, jotka sisältävät monimuotoisia luonnollisia kohinakuvioita. Verkko oppii kilpailullisella koulutuksella tuottamaan HR-kuvista LR-kuvia, jotka kärsivät luonnollisen tyyliä huonontumisesta, mitä esiintyy esimerkiksi huonoista valvontakameroista. Näin voidaan saada tarpeellinen määrä HR- ja LR-kuvia pareiksi superresoluutioverkon koulutukseen.

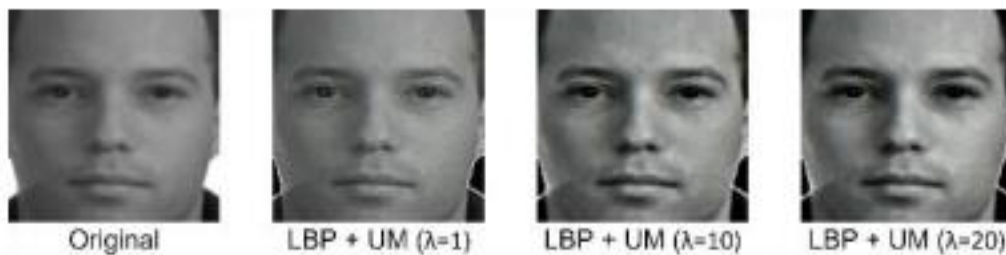
4.4 Esikäsittelyn hyöty

Monien eri SR-menetelmien ongelmana on niiden kyky säilyttää kasvojen yleinen rakenne, vaikka ne tuottavat luonnollisen näköisiä kuvia. Kun tuotettuja kuvia koitetaan käyttää kasvojentunnistuksessa, on niiden tarkkuus alhainen, koska ne tuottavat kuvan kasvoille uusia yksityiskohtia säilyttämättä alkuperäisiä muotoja (Shahbakhsh, Hassanpour. 2020). Ennen superresoluution tuottamista syötekuvaa voidaan esikäsitellä, jotta tulos olisi alkuperäisen kaltainen. Reunainformaatiolla on tärkeä rooli säilyttämään kas-

vojen rakenne ja saamaan kuvasta korkeataajuinen tieto, joten LR-kuvan reunoja on käytettävä oikealla tavalla. Shahbakhsh ja Hassanpour esittävät paperissaan menetelmän säilyttämään kasvojen muodot.

1. Kasvojen muotoja voidaan ottaa talteen käyttämällä *paikallista binäärikuviota* (Local Binary Patternia, LBP) ja ne lisätään syötekuvaan.
2. Talteen otetut reunat terävöitetään käyttämällä *epäterävää maskia* (Unsharpened Mask, UM).

Tämä esikäsittely voi auttaa esitettyä mallia paremmin säilyttämään kasvojen yksityiskohdat. Kuvassa 3 on esitetty LBP:n ja UM:n vaikutus kuvaan. λ on UM-filtterin vahvuustekijä. Shahbakhsh ja Hassanpour käyttävät verkossaan myös häviöfunktiota, jonka tavoitteena on säilyttää reunainformaatio.



Kuva 3 – Esikäsittelyn vaikutus syötekuvaan. (Shahbakhsh, Hassanpour. 2020)

4.5 Laadunmittaamisen kehittäminen

Kaikki tähän asti ehdotetut tavat mitata eri superresoluutioiden laatua omaavat erilaisia etuja ja haittoja. Jotta koneoppimista voidaan ohjata parempaan suuntaan, täytyy niiden kykyä tuottaa halutunlaista tulostetta pystyä mittaamaan nykyistä paremmin. Ilman hyviä mittaamenetelmiä, tällä hetkellä tutkija voi esimerkiksi optimoida mallinsa tuottamaan parhaan PSNR-arvon, koska se on nykyään yksi käytetyimmistä mittaamenetelmistä. Tämä kuitenkin yleensä johtaa liian pehmeisiin kuviin, joista puuttuvat tarkat yksityiskohdat, eivätkä tuotetut kuvat ihmisten mielestä olisi hyviä. Parempia mittaustapoja on kehitteillä ja käytössä, mutta ihmisen havaitsemaa laatua mittaavia malleja ei ole yleisesti hyväksytyssä käytössä. Ihmisten käyttämistä tuotettujen kuvien arviointiin on tehty, ja ne tuottavat tietysti hyvää tietoa kuvien visuaalisesta laadusta. Kuitenkaan tuloksia ei voida toistaa uudelleen ja ihmiset omaavat omia taipumuksiaan arvioinnissa, riippuen tilanteesta ja arviointiasteikoista. Paremmat mittaustavat tarvitaan, jotta tutkijat voivat olla varmoja, että muutokset malleihin tuottavat parempaa tulosta. Varsinkin kasvojen superresoluutiassa luonnollisemman näköiset tulokset ovat tärkeämpi metriikka, kuin vaikka PSNR, johtuen ihmisten kyvystä huomata outoudet kasvoista helposti. Myös sokeille

IQA-metodeille on tarvetta, jotta kuvanlaatua voidaan mitata vertaamalla alkuperäisiin HR-kuviin (Wang et al., 2021).

5 Keskustelu

GAN-verkot ovat hyvin mielenkiintoinen tutkimuskohde, joilla voidaan tuottaa paljon erilaista materiaalia, mikä ei rajoitu vain tämän paperin superresoluutioihin. On kiinnostavaa nähdä, mitä kaikkea tulevaisuudessa näillä voidaan pystyä tuottamaan. Tälläkin hetkellä netin palveluita käyttäen voi näitä hyödyntävillä verkoilla kuka vain tuottaa SR-kuvia tai täysin uusia kuvia tekstikuvauksen pohjalta.

Tulevaisuudessa saa myös nähdä, että mitkä ratkaisut osoittautuvat parhaimmiksi superresoluutioiden tuottamiseen, koska tällä hetkellä eri tutkimukset ehdottavat monia erilaisia parannuksia GAN-verkkojen toimintaan. Kuten Yang Zhang ja muut paperissaan esittivät, ei eri ratkaisuja voi helposti yhdistää ilman sivuvaikutuksia tuloksiin. Erilaisten ratkaisujen käyttö lisää myös prosessoinnin tarvetta superresoluutioiden tuottamiseen, mikä rajoittaa niitten käyttöä, varsinkin reaaliaikaisissa käyttökohteissa.

Vaikka useita hyvin suoriutuvia superresoluution malleja on esitetty eri tutkimuspapereissa, ovat niiden tulokset oikean maailman käyttökohteissa usein huonompia. Mallien kouluttaminen tietojoukoilla, joiden laatua on huononnettu manuaalisesti säädetyillä tavoilla, mikä ei vastaa oikeassa elämässä tapahtuvaa degradaatiota, on usein syyppää tähän. Tämä on ongelma, mikä tarvitsee nopeaa ratkaisua. Esitetyt ratkaisut kärsivät mallien opettamisen vaikeudesta ja liian täydellisistä oletuksista (Wang, 2020).

Monet tutkimuspaperit tutkivat SR-tekniikoita kasvojentunnistusta varten, mutta näistäkin käyttökohteista puuttuu osoitukset toimivuudesta reaali maailman käyttökohteissa. Testidatasta saadut lupaavat tulokset eivät välttämättä toteudu todellisuudessa. Vaikka LR-kasvokuvasta voidaan saada oikean näköinen suurennos, on myös pidettävä mielessä, että SR-kuva on vain GAN-verkon arvio, siitä miltä kasvot voisivat näyttää. SR-kuvia ei varmastikaan voida käyttää todisteina rikoksiin. Näen kuitenkin, että niitä voidaan hyödyntää esimerkiksi etsintäkuulutuksissa samalla tavalla, miten poliisin kasvopiirroksia käytetään epäiltyjen tunnistukseen. Jos SR-kuvia aletaan käyttää automatisoituihin kasvojentunnistusjärjestelmiin, voisivat ne tuottaa paljon vääriä osumia.

Verkkojen prosessoinnin tarpeen määrä olisi ollut yksi osa-alue, jota voisi tutkia vielä tarkemmin. Useat lähteet puhuvat prosessoinnin tarpeen lisääntymisestä verkon monimutkaistuesssa, jotta voidaan tuottaa parempia tuloksia. Lähteet eivät kuitenkaan mainitse

tarkkoja lukuja, joten on vaikea arvioida kuinka suuria määriä tehoa vaaditaan superresoluution tuottamiseen. Reaaliaikaiselle superresoluutiolle on paljon tarvetta, joten on hyvä tietää minkälaiset verkot sopivat tähän käyttöön. Esimerkiksi Nvidian näytönohjaimissa on niiden käyttämien syväoppimisen teknologioita varten tarkoitukseen rakennettuja prosessointiytimiä (Nvidia, 2022). Tällaisten kehittäminen ja rakentaminen yleiseen käyttöön, esimerkiksi kuluttajille suunnatuissa televisioissa, on kallista.

6 Yhteenveto

GAN-verkot ovat alkaneet olla tutkijoille kiinnostavimpia neuroverkkoja luonnollisten tuloksien saamiseen, mikä on erityisen tärkeää superresoluution kohteissa. Nykyisin voidaan tuottaa jo vakuuttavia tuloksia, mutta varsinkin kasvojen superresoluutiossa on vielä haasteita, joille ei ole löydetty parhaimpia ratkaisuja. Erilaiset ratkaisut keskittyvät tiettyyn osaongelmaan, eikä ratkaisuja ole helppo yhdistää yhdeksi toimivaksi verkoksi. Verkon monimutkaistuessaa lisääntyvät muistin ja prosessointitehon tarve, mikä laskee ratkaisujen käyttöä reaaliaikaisiin käyttökohteisiin, kuten pakatun suoratoistokuvan parantamiseen TV:ssä. Tästä huolimatta, kun halutaan luonnolliselta näyttäviä tuloksia, ovat GAN-verkot osoittautuneet yksiksi parhaimmiksi syväoppimisen tavoiksi tuottaa superresoluutioita. Superresoluutiota tuottaessa usein halutaan ihmiselle miellyttävän näköistä tulostetta, mutta esimerkiksi lääketieteen kuvantamisen superresoluutiossa tämä ei ole tärkeää ja siinä käytetään muita kuin GAN-verkkoon perustuvia SR-menetelmiä.

Nykyään esimerkiksi lääketieteessä ja kaukokartoituksessa voidaan hyödyntää superresoluutiota, mutta turvallisuussektorin kasvojentunnistuksessa superresoluution käytössä on paljon lisähaasteita, mitä vielä koitetaan ratkaista. Esimerkkejä haasteista ovat kattavan koulutusdatan hankkiminen ja koulutusdatan muuttaminen huonolaatuiseksi realistisella tavalla. Tutkimukset ovat saaneet lupaavia tuloksia testeissä, mutta oikeassa maailmassa tulokset eivät ole niin hyviä. Tavalliset kuluttajat voivat jo kokeilla superresoluutiota esimerkiksi netin kautta toimivilla palveluilla tai videopeleissä hyödyntämällä näytönohjainvalmistajien kehittämiä menetelmiä.

Superresoluutioiden laatua mitataan eri kvalitatiivisilla ja subjektiivisilla menetelmillä, joissa kaikissa on jotain kehittämisen tarvetta. Objektiiiset menetelmät voidaan laskea aritmeettisesti, mutta ne eivät kuvasta ihmisten kokemusta SR-kuvien laadusta. Subjektiiviset menetelmät kärsivät varianssista ja toistettavuuden haasteista. Minkä mittarin tärkeyttä korostetaan laadunmittauksessa, on vielä avoin kysymys. Kuvanlaadunmittarit

ovat tärkeitä, jotta tutkijat voivat mitata verkon kehittymistä ja verrata tuloksia muihin verkkoihin.

Tulevaisuudessa GAN-verkot ja superresoluutio tulevat varmasti kiinnostamaan tutkijoita ja tulemme näkemään uusia käyttökohteita, enemmän reaaliaikaisia malleja sekä parempia verkkoja eri tarkoituksiin. Tärkeimmiksi ongelmiksi ratkaistavaksi jäävät vielä harjoitusdatan muuttaminen vastaamaan oikean maailman LR-kuvia ja parempien kuvanlaatumittareiden määrittäminen mittaamaan uudelleenrakentamisen laatua SR-kuvissa.

Lähdeluettelo

A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," in IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

D. Lepcha, B. Goyal, A. Dogra, V. Goyal, Image super-resolution: A comprehensive review, recent trends, challenges and applications, Information Fusion, Volume 91, 2023, Pages 230-260, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2022.10.007>.

K. Singla, R. Pandey, U. Ghanekar, A review on Single Image Super Resolution techniques using generative adversarial network, Optik, Volume 266, 2022, 169607, ISSN 0030-4026, <https://doi.org/10.1016/j.ijleo.2022.169607>.

M. B. Shahbakhsh, & H. Hassanpour, (2022). Empowering face recognition methods using a GAN-based single image super-resolution network. International Journal of Engineering, Transactions A, Basics, 35(10), 1858.
doi:<https://doi.org/10.5829/IJE.2022.35.10A.05>

Zhao, S., Sun, J., Ou, H., & Lin, Y. (2022). A novel multi-task face super-resolution framework embedding degraded augmented GAN networks. Journal of Physics: Conference Series, 2303(1), 012061. doi:<https://doi.org/10.1088/1742-6596/2303/1/012061>

Y. Zhang, I. W. Tsang, J. Li, P. Liu, X. Lu and X. Yu, "Face Hallucination With Finishing Touches," in IEEE Transactions on Image Processing, vol. 30, pp. 1728-1743, 2021, doi: 10.1109/TIP.2020.3046918.

Z. Wang, J. Chen and S. C. H. Hoi, "Deep Learning for Image Super-Resolution: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3365-3387, 1 Oct. 2021, doi: 10.1109/TPAMI.2020.2982166.

S. Kumar and S. Dhawan, "A Detailed Study on Generative Adversarial Networks," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 641-645, doi: 10.1109/ICCES48766.2020.9137883.

K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng and F. -Y. Wang, "Generative adversarial networks: introduction and outlook," in IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588-598, 2017, doi: 10.1109/JAS.2017.7510583.

C. -C. Hsu, C. -W. Lin, W. -T. Su and G. Cheung, "SiGAN: Siamese Generative Adversarial Network for Identity-Preserving Face Hallucination," in IEEE Transactions on Image Processing, vol. 28, no. 12, pp. 6225-6236, Dec. 2019, doi: 10.1109/TIP.2019.2924554.

J. Hall, M. G. Bocanegra and R. J. Haddad, "Optimizing Generative Adversarial Networks for Low-Resolution Image Enhancement," 2020 SoutheastCon, 2020, pp. 1-6, doi: 10.1109/SoutheastCon44009.2020.9368265.

M. Cao, Z. Liu, X. Huang and Z. Shen, "Research for Face Image Super-Resolution Reconstruction Based on Wavelet Transform and SRGAN," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 448-451, doi: 10.1109/IAEAC50856.2021.9390748.

Nvidia, DLSS (Deep learning Super Sampling) -teknologia. Tuotteen esittelysivu yrityksen sivustolla, sivu ladattu 5.12.2022. <https://www.nvidia.com/fi-fi/geforce/technologies/dlss/>