



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
STUDIJŲ PROGRAMA: INFORMATIKA

Generatyviniai besivaržantys neuroniniai tinklai veidams generuoti
Generative adversarial neural networks for face generation

Baigiamasis magistro darbas

Atliko: Dovydas Lizdenis

VU el. p.: dovydas.lizdenis@mif.stud.vu.lt

Vadovas: Prof. Dr. Aistis Raudys

Recenzentas: Prof. Dr. Olga Kurasova

Vilnius – 2024

Santrauka

Viena iš populiariausių šių laikų technologijų yra dirbtiniai neuroniniai tinklai. Ši technologija yra plačiai taikoma medicinoje, policijoje, IT srityse. Neuroninių tinklų pagrindinė mintis yra apmokyti algoritmą pagal suteiktus duomenis, kad atlikti tam tikras užduotis neprogramuojant specifinių veiksmų, galima sakyti, kad neuroniniai tinklai galėtų sugalvoti atsakymus į tam tikras pateiktas užduotis. Tačiau vienas iš išskirtinių žmogaus gebėjimų, kuris yra sunkesnis įgyvendinti kompiuteriuose, yra mokėjimas sukurti kažką naujo ir realaus. Žmogaus mokėjimas įsivaizduoti įvairius pasaulius, aplinkas, žmones, vaizdus ir juos perteikti į knygas, muziką, paveikslus, taip sukuriant naujus realius dalykus. Kad atkartoti tokį gebėjimą pasinaudojant dirbtiniu intelektu kompiuteriuose buvo sukurti įvairūs algoritmai įskaitant generatyvinius besivaržančius tinklus (GAN), kurie ir bus tiriami šiame darbe, nes dabar jie yra vieni iš populiariausių algoritmų naudojamų nuotraukų generavimui. Be to GAN, gali atkartoti žmogaus gebėjimą kurti naujus ir tikroviškus dalykus tokius, kaip žmonių nuotraukas, gali sujungti dvi nuotraukas ar sugeneruoti naują unikalią nuotrauką, užbaigti nuotraukos dalį, generuoti nuotrauką iš teksto, padidinti nuotraukos raišką ir t.t. Todėl šiame tyrime bus nagrinėjama, kur veidų generavimas naudojamas, kaip veikia originalus GAN, kokie GAN dabar yra sukurti veidams generuoti, sujungti ar pakeisti vaizdo įrašuose. Be to bus apžvelgtas rastų GAN veikimas ir bus palyginti jų generuojami rezultatai tarp numatytų kategorijų.

Summary

One of the most popular technologies today is artificial neural networks. This technology is widely used in medicine, policing and IT fields. The main idea behind neural networks is to train an algorithm on the basis of given data to perform certain tasks without programming specific actions, so that neural networks can be said to be able to come up with answers to certain given tasks. However, one of the unique human abilities that is more difficult to implement in computers is the ability to create something new and real. The human ability to imagine different worlds, environments, people, images, and to translate this into books, music, paintings, thus creating real new things. To replicate this ability using artificial intelligence in computers, various algorithms have been developed, including generative adversarial networks (GANs), which will be investigated in this thesis, as they are now one of the most popular algorithms used for photo generalization. In addition GANs can replicate the human ability to create new and realistic things such as photos of people, can merge two photos or generate a new unique photo, complete a part of a photo, generate a photo from text, increase the resolution of a photo, etc. This study will therefore look at where face generation is used, how the original GAN works, and what GANs are now developed to generate, merge or modify faces in videos. In addition, the performance of the GANs found will be reviewed and the results they generate will be compared between the intended categories.

Turinys

Įvadas.....	8
1. Generatyvinių besivaržančių neuroninių tinklų apžvalga ir taikymo sritys.....	12
1.1. Originalaus GAN apžvalga.....	12
1.1.1 GAN veikimo principas.....	12
1.1.2 GAN diskriminatoriaus veikimas.....	14
1.1.3 GAN generatoriaus veikimas.....	14
1.1.4 GAN nuostolio funkcija.....	16
1.1.5 GAN trūkumai ir problemos.....	16
1.2. Veido generavimo panaudojimo būdai.....	17
1.3. Veido generavimo iš kelių nuotraukų panaudojimo būdai.....	18
1.4. Veido generavimo arba keitimo panaudojimo būdai vaizdo įrašė.....	20
1.5. Taikymo sričių ir originalaus GAN apibendrinimas.....	22
2. Generatyvinių besivaržančių neuroninių tinklų analizė.....	25
2.1. Veido nuotraukas generuojančių GAN analizė.....	25
2.1.1 StyleGAN2 analizė.....	25
2.1.2 BigGAN analizė.....	27
2.1.3 PG-GAN (ProGan) analizė.....	28
2.1.4 Apibendrinimas.....	30
2.2. GAN skirtų veidų generavimui iš kelių nuotraukų analizė.....	30
2.2.1 InterFaceGAN analizė.....	30
2.2.2 BlendGAN analizė.....	33
2.2.3 StyleGAN analizė.....	35
2.2.4 Apibendrinimas.....	36
2.3. GAN skirtų veidų keitimui vaizdo įrašuose analizė.....	37
2.3.1 FSGAN analizė.....	37
2.3.2 DeepFaceLab analizė.....	40
2.3.3 Apibendrinimas.....	43
2.4. Analizės apibendrinimas.....	44
3. Generatyvinių besivaržančių neuroninių tinklų eksperimentinė patikra ir vertinimas.....	45
3.1. GAN vertinimo metodai IS, FID ir SSIM.....	45

3.2. GAN įvertinimų palyginimai.....	47
3.2.1 GAN įvertinimas pagal veidų generavimą.....	47
3.2.2 GAN įvertinimas pagal veidų generavimą iš kelių nuotraukų.....	50
3.2.3 GAN veidų generavimo vaizdo įrašuose palyginimas.....	53
3.3. Eksperimentinės dalies apibendrinimas.....	55
Išvados.....	58
Literatūra.....	60
Priedai.....	63
1 priedas. IEEE Access pateiktas straipsnis.....	63

Paveikslų ir lentelių sąrašas

Lentelės

1 lentelė. GAN palyginimas, kurie generuoja veidus pagal IS ir FID.....	47
2 lentelė. GAN, kurie generuoja veidus iš kelių veidų, palyginimas pagal FID.....	50
3 lentelė. GAN, kurie generuoja veidus vaizdo įrašuose, palyginimas pagal SSIM.....	53

Paveikslai

1 pav. DRAW principo schema [GDG+15].....	9
2 pav. StyleGAN sugeneruota žmogaus nuotrauka iš https://thispersondoesnotexist.com	9
3 pav. GAN mokymosi principas [GOO22].....	13
4 pav. GAN principo schema.....	13
5 pav. Veido pakeitimas vaizdo įrašė. Šaltinis: spectrum.ieee.org	22
6 pav. StyleGAN generatoriaus architektūra [KLA19].....	26
7 pav. BigGAN sugeneruotos nuotraukos Tensorflow.org.....	27
8 pav. PG-GAN mokymo schema [KAL+18].....	29
9 pav. InterFaceGAN bendras nuotraukos manipuliavimo pavyzdys [SGT+20].....	31
10 pav. InterFaceGAN sąlyginio manipuliavimo pavyzdys [SGT+20].....	32
11 pav. InterFaceGAN rezultatų pavyzdžiai [SGT+20].....	32
12 pav. BlendGAN veikimo pavyzdys [MQZ+21].....	34
13 pav. BlendGAN rezultatų pavyzdžiai [MQZ+21].....	35
14 pav. StyleGAN veidų sujungimo pavyzdžiai [KLA19].....	36
15 pav. FSGAN veikimo pavyzdys [NKH19].....	38
16 pav. FSGAN išvaizdos žemėlapis [NKH19].....	39
17 pav. FSGAN sugeneruoti pavyzdžiai [NKH19].....	40
18 pav. DeepFaceLab veido išgavimo pavyzdys [PGC+21].....	41
19 pav. DeepFaceLab mokymosi pavyzdys [PGC+21].....	42
20 pav. DeepFaceLab veido generavimo pavyzdys [PGC+21].....	42
21 pav. DeepFaceLab veido generavimo pavyzdys [PGC+21].....	43
22 pav. IS pasiskirstymo pavyzdys medium.com.....	46
23 pav. BigGAN gauti rezultatai.....	48
24 pav. PG-GAN gauti rezultatai.....	49
25 pav. StyleGAN2 gauti rezultatai.....	49

26 pav. StyleGAN gauti rezultatai.....	51
27 pav. InterFaceGAN gauti rezultatai.....	52
28 pav. BlendGAN gauti rezultatai.....	52
29 pav. DeepFaceLab ir FSGAN gauti rezultatai.....	54
30 pav. StyleGAN atrinkti realistiški veidai.....	55
31 pav. StyleGAN prastai sugeneruoti veidai.....	56
32 pav. InterFaceGAN prastai sugeneruoti veidai.....	56
33 pav. DeepFaceLab prastai sugeneruotas kadras.....	57

Įvadas

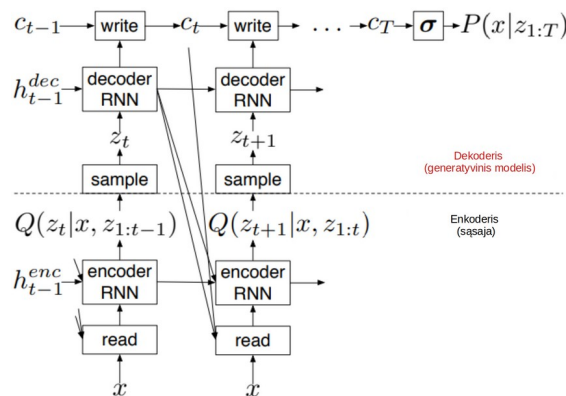
Temos aktualumas – viena iš populiariausių šių laikų technologijų yra dirbtiniai neuroniniai tinklai. Ši technologija yra plačiai taikoma medicinoje, policijoje, IT srityse. Neuroninių tinklų pagrindinė mintis yra apmokyti algoritmą pagal suteiktus duomenis, kad atlikti tam tikras užduotis. Neuroniniai tinklai gali su dideliu tikslumu atpažinti objektus, žmones nuotraukose arba vaizdo įrašuose, gali greitai ir tiksliai suklasifikuoti nuotraukas, atpažinti žmonių veidus bei atpažinus veidą gali priskirti žmogų, kurio veidas yra atpažintas, nustatyti lytį ar net amžių [BWL20][DR17][KRR+10]. Taip pat neuroniniai tinklai gali atsižvelgiant į praeitus duomenis prognozuoti pasikeitimus ateityje. Neuroninių tinklų panaudojimas yra labai lankstus ir juos galima pritaikyti įvairiausiose srityse. Praktiškai neuroninius tinklus galima pritaikyti ten, kur yra įmanoma kompiuterį apmokyti daryti užduotis.

Tačiau viena iš išskirtinių žmogaus savybių yra gebėjimas sukurti kažką naujo ir realaus. Žmogaus mokėjimas įsivaizduoti įvairius pasaulius, aplinkas, žmones, vaizdus ir juos perteikti į knygas, muziką, paveikslus, taip sukuriant naujus realius dalykus. Tam, kad atkartoti tokį gebėjimą pasinaudojant dirbtiniu intelektu kompiuteriuose yra naudojami skirtingi algoritmai tokie, kaip: konvoliuciniai neuroniniai tinklai (*angl. Convolutional neural networks, CNN*) [DS15], nuolatinis dėmesingas rašytojas (*angl. Deep Recurrent Attentive Writer, DRAW*) [GDG+15], generatyviniai besivaržantys tinklai (*angl. Generative adversarial networks, GAN*) [GAM+14].

Pradėkime nuo CNN algoritmo, kuriuo pasinaudojant galima sugeneruoti vaizdą. Pavyzdys, kuris bus nagrinėjimas yra kėdžių vaizdų generavimas [DS15]. Pirmiausia reikia apmokyti neuroninį tinklą, kuriam yra pateikiami jau minėti kėdžių vaizdai. Po apmokymo pateikus tam tikrus parametrus: kėdės tipą, atvaizdavimą, spalvą, neuroninis tinklas gali sugeneruoti kėdės vaizdą. Gauta vaizdo dydis būna nuo 48x48px iki 128x128px, kurį būtų galima padidinti, tačiau tai reikalauja daugiau kompiuterinių resursų [DS15].

Pagrindinė DRAW algoritmo mintis yra pasinaudojant koduotoju ir dekoduoju, kurie yra pasikartojantys tinklai (*angl. Recurrent neural network, RNN*), nuskaityti gautą įvestį dalimis atsižvelgiant į jau atliktus žingsnius ir gautą informaciją užkoduoti bei išsaugoti. Po to dekoduoti ir rekonstruoti įvestį iš naujo. Be to nuskaitant duomenis koduotojo tinklas žino prieš tai išvestą dekoduoju rezultata, pagal kurį gali rasti informaciją, kurią reikės nuskaityti toliau. Po duomenų nuskaitymo, DRAW algoritmas gali sugeneruoti vaizdus iš turimų išsaugotų ir

užkoduotų duomenų, sukuriant naujus vaizdus ar objektus [GDG+15]. DRAW algoritmo schema matoma 1 pav.



1 pav. DRAW principo schema [GDG+15]

Paveiksle yra matoma DRAW algoritmo schema. Joje parodyti visi žingsniai, kurie buvo aprašyti. Apačioje po brūkšniuota linija yra nuskaitymo ir kodavimo žingsniai, iš jų pereinama į „sample“ žingsnį (žr. 1 pav.), kuriame yra išsaugoma informacija ir vėliau iš šio žingsnio yra generuojami vaizdai. Virš brūkšniuotos linijos yra dekodavimo ir įrašymo žingsniai. Be to nuotraukoje matosi rodyklės, kurios nurodo, kaip tinklai dalinasi informacija tarpusavyje po kiekvieno nuskaitymo.

Paskutinis aprašomas algoritmas yra GAN. Šis algoritmas dabar yra vienas iš populiariausių naudoti nuotraukų sintetizavimui (*angl. image synthesis*). Nuotraukų sintetizavimas yra gan svarbus ir jį galima pritaikyti meno generavimui, kompiuterio dizainui, nuotraukų redagavimui, virtualiai realybei ir t.t. [FHR+21] Be to naudojant GAN algoritmą, galima sukurti aukštos kokybės žmonių veidų nuotraukas 2 pav. [KLA+20].



2 pav. StyleGAN sugeneruota žmogaus nuotrauka iš <https://thispersondoesnotexist.com>

Paveiksle matoma nuotrauka sukurta naudojant StyleGAN2 algoritmu.

Taip pat naudojant GAN galima sukurti vaizdo įrašus su kitų žmonių veidais [WES19], generuoti iš nuotraukos į nuotrauką [EAD+20], generuoti realistišką tekstą [ZGF+17], nuotraukose padidinti raišką, generuoti iš teksto į nuotrauką, neturint dalies nuotraukos ją užbaigti ar papildyti (*angl. image in-painting*) ir dar daugiau [FHR+21]. GAN bendrasis principas yra pateiktas 4 pav. Nuotraukoje yra pavaizduota GAN principinė schema, kurioje matosi, du besivaržantys tarpusavyje tinklai – generatorius ir diskriminatorius. Generatorius gauna atsitiktinę įvestį iš kurio toliau generuoja nuotrauką ir gautą rezultatą pateikia diskriminatoriui. Vienu metu diskriminatorius gauna dvi nuotraukas – vieną sugeneruotą generatoriaus, o kitą tikrą. Turint dvi nuotraukas diskriminatorius sprendžia ar generatoriaus sugeneruota nuotrauka yra reali. Jei nuotrauką diskriminatorius priskiria prie tikrų, tuomet ši nuotrauka yra pateikiama, kaip rezultatas. O jei nuotrauka priskiriama prie netikrų, generatorius iš to mokosi ir generuoja naują nuotrauką, kol įtikina diskriminatorių, kad tai yra reali nuotrauka [GAM+14].

Šiame darbe bus tiriami skirtingi GAN algoritmai, kurie generuoja veidus, nes dabar nėra išsamių lyginamųjų darbų, kurie aptaria ir palygina dabartinius algoritmus veidams generuoti. Be to GAN, kaip buvo minėta anksčiau, gali atkartoti žmogaus gebėjimą kurti naujus ir tikroviškus dalykus, skirtingai nuo DRAW ar CNN. Todėl šiame darbe visų pirma bus apžvelgta, kur veidų generavimas yra naudojamas ir tuomet bus apžvelgiami, kokie GAN dabar yra sukurti veidų generavimui pagal skirtingas kategorijas: veido generavimas, veido generavimas iš kelių veidų (sujungimas), bei veido generavimas arba keitimas vaizdo įrašė. Toliau bus nagrinėjamas pačio pirmojo GAN principo veikimas pagal Ian Goodfellow. Galiausiai kiekvienoje kategorijoje rasti algoritmai bus apžvelgti ir sulyginti.

Darbo tikslas – atlikti tyrimą apie veidus generuojančius generatyvinius besivaržančius tinklus ir palyginti jų generuojamus rezultatus pagal šias kategorijas: veidų generavimas, veidų generavimas sujungiant kelis veidus ir veidų generavimas arba veido keitimas vaizdo įrašuose.

Uždaviniai:

1. Apžvelgti GAN taikymo sritis ir generatyvinius besivaržančius neuroninius tinklus šiose kategorijose:
 - Veidų generavime;
 - Veido generavime iš kelių veidų ar nuotraukų;
 - Veido keitime vaizdo įrašuose.

2. Atlikti šiose kategorijose taikomų GAN modelių analizę ir nustatyti jų privalumus bei trūkumus;
3. Eksperimentiškai patikrinti ir įvertinti GAN modelių rezultatus, taikant skirtingas vertinimo metodikas.

Numatomi rezultatai: bus nustatyta kuriose srityse veido generavimas yra naudojamas ir bus apžvelgti GAN, kurie generuoja veidus, generuoja veidą iš kelių veidų ir keičia veidą vaizdo įrašė. Taip pat bus išanalizuotas ir išsiaiškintas originalaus GAN veikimas, kad toliau būtų galima atlikta išsamią išsirinktų GAN analizę, kas buvo pakeista norint pagerinti generuojamus rezultatus. Galiausiai bus atlikta eksperimentinė analizė ir palyginti generatyviniai besivaržantys neuroniniai tinklai tarpusavyje pagal minėtas kategorijas.

1. Generatyvinių besivaržančių neuroninių tinklų apžvalga ir taikymo sritys

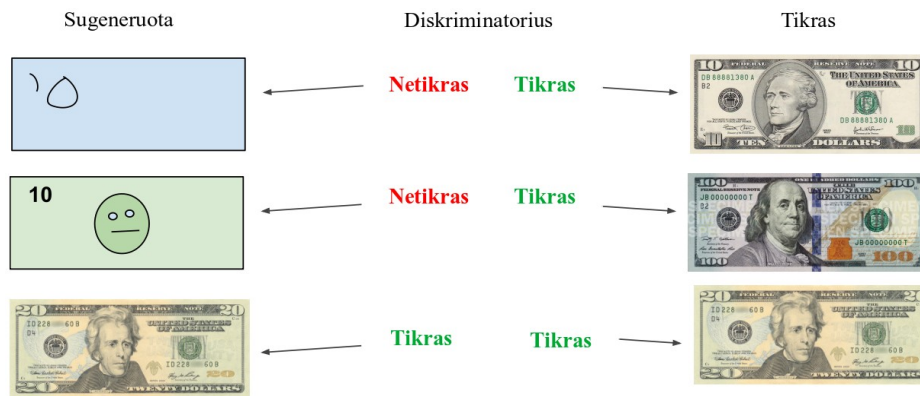
Veidų generavimas tapo populiariesnis ir labiau naudojimas po GAN pristatymo, nes GAN naudojimas atvėrė daug naujų krypčių. Naudojant įprastus mašininio mokymo metodus, norint sugeneruoti kokią nors nuotrauką arba veidą, reikia labai daug mokymo duomenų. Naudojant GAN, duomenų kiekis gali būti mažesnis, norint išgauti panašų rezultatą lyginant su įprastu mašininio mokymu. Be to GAN realistiškus veidus gali sugeneruoti iš „triukšmo“ (*angl. noise*) [KST+22]. Dabar nėra sunku šiek tiek paieškoti ir rasti internete puslapių, teikiančių tokias paslaugas, kaip: žmogaus veido generavimo, vaizdo įrašų su norimu tekstu kūrimo ar žmogaus veidų kūrimo, sujungiant kelių žmonių veidus ir panašiai. Todėl šiame skyriuje bus aprašoma, kur veido generavimas yra naudojamas ir bus aptarti veido generavimo keliami pavojai bei apžvelgtas originalus GAN.

1.1. Originalaus GAN apžvalga

Šiame poskyryje bus detaliau aprašoma, kas yra generatyviniai besivaržantys neuroniniai tinklai (GAN), kaip jie veikia bei, kaip jie mokosi varžydamiesi tarpusavyje.

1.1.1 GAN veikimo principas

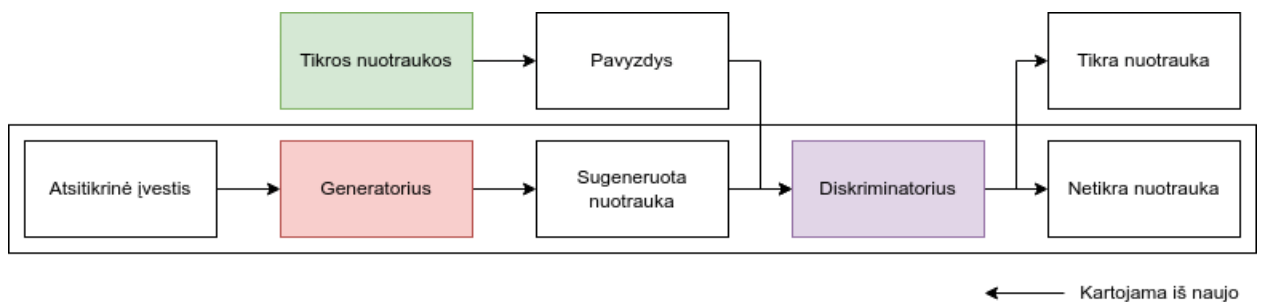
GAN modelis buvo pateiktas 2014 metais Ian Goodfellow ir kt [GAM+14]. Pačio GAN modelio principas yra dviejų neuroninių tinklų tarpusavio varžymasis. Visą mokymosi procesą galima įvardinti, kaip pastovų dviejų žaidėjų žaidimą, kur vienas iš žaidėjų yra generatorius, o kitas žaidėjas yra diskriminatorius. Generatoriaus tikslas yra sugeneruoti naujus ir įtikinamus duomenis iš jam pateikto duomenų rinkinio, o diskriminatoriaus tikslas yra atskirti netikrus duomenis, kuriuos pateiks generatorius. Jeigu pateikti duomenys iš generatoriaus buvo nustatyti kaip netikri, tada generatorius yra nubaudžiamas, kad pateikė netikrus duomenis ir taip mokosi. Jeigu generatorius sugeneruoja įtikinamus duomenis, tai ir bus galutinis GAN mokymosi proceso rezultatas [GAM+14][GOO22]. Bendroji mokymosi schema pateikiama 3 pav., nusakanti kaip vyksta procesas.



3 pav. GAN mokymosi principas [GOO22]

Paveiksle matomas bendro GAN modelio mokymosi principas, kuriame pavaizduotas generatoriaus progresas, diskriminatoriaus vaidmuo ir kaip lyginami duomenys. Pradžioje aiškiai matosi, kad generatoriaus pateikti duomenys yra netikri ir neatitinkantys realių vaizdų. Po nepavykusio bandymo generatorius sugeneruoja šiek tiek geresnį rezultatą ir vėl pateikia diskriminatoriui įvertinti, nors antrasis bandymas vizualiai geresnis, tačiau diskriminatorius vaizdą įvertina kaip netikrą. Paskutiniame žingsnyje matosi, kad generatorius sukuria praktiškai vienodą, šiuo atveju banknotą, kuris yra pripažintas realiu ir tai yra galutnis rezultatas. Be to mokymosi procesas yra pakaitinis, kuriuo metu mokosi arba generatorius arba diskriminatorius.

Bendra GAN veikimo principo schema pateikiama 4 pav.



4 pav. GAN principo schema

Paveiksle matoma, kaip GAN modelis veikia iš vidaus. Viskas prasideda nuo atsitiktinės įvesties į generatorių, toliau generatorius sugeneruoja nuotrauką ir pateikia į diskriminatorių. Kuris gauna dvi įvestis – tikrą nuotrauką ir sugeneruotą. Diskriminatorius iš nurodytų nuotraukų nusprendžia ar generatoriaus sugeneruota nuotrauka yra tikra, jei taip generatorius laimi ir gaunamas atsakymas yra generatoriaus sugeneruota nuotrauka. Kitu atveju nėra gaunamo jokio rezultato ir generatorius yra nubaudžiamas. Nubaudus generatorių, jis iš to pasimoko ir vėl ciklas yra kartojamas [GAM+14].

1.1.2 GAN diskriminatoriaus veikimas

GAN diskriminatorius yra klasifikatorius, kuris taip pat yra daugiasluoksnis perceptronas. Jis bando atskirti tikrus duomenis nuo generatoriaus sukurtų duomenų. Diskriminatorius gali naudoti bet kokią tinklo architektūrą, atitinkančią klasifikuojamų duomenų tipą. Diskriminatorius treniruodamasis gauna dviejų tipų duomenis:

- Tikri duomenų atvejai, pvz., tikros žmonių nuotraukos. Diskriminatorius naudoja šiuos atvejus kaip tinkamus pavyzdžius mokymosi metu;
- Generatoriaus sukurti netikri duomenų egzemplioriai. Diskriminatorius naudoja šiuos atvejus kaip netinkamus pavyzdžius mokymosi metu.

Schema pateiktoje 4 pav. galima matyti dviejų tipų duomenis, patenkančius į diskriminatorių. Diskriminatorius jungiasi prie dviejų praradimo funkcijų. Treniruojant diskriminatorių yra naudojamas diskriminatoriaus nuostolis ir ignoruojamas generatoriaus nuostolis. Diskriminatorius mokymosi metu yra atliekami šie veiksmai:

- Diskriminatorius klasifikuoja tiek tikrus, tiek netikrus generatoriaus duomenis;
- Dėl diskriminatoriaus klaidų jis yra baudžiamas už tikrojo atvejo klasifikavimą kaip netikro arba netikro atvejo klasifikavimą, kaip tikro;
- Diskriminatorius atnaujina savo svorius tinkle, atsižvelgiant į gautą praradimo rezultatą.

Be to diskriminatorius mokosi tik tada, kai generatorius nesimoko, nes kitu atveju generatoriaus mokymas būtų sudėtingesnis [GAM+14][GOO22].

1.1.3 GAN generatoriaus veikimas

Kaip ir buvo minėta GAN generatorius mokosi kurti netikrus duomenis įtraukdama grįžtamąjį ryšį iš diskriminatoriaus, jei nepavyksta jo apgauti. Generatorius išmoksta apgauti diskriminatorių tuomet, kai diskriminatorius sugeneruotą vaizdą suklasifikuoja, kaip tikrą. Generatoriaus mokymas reikalauja gilesnio bendravimo tarp generatoriaus ir diskriminatoriaus, nei reikalauja diskriminatorius mokydamasis. Generatorius mokosi atlikdamas šiuos žingsnius:

- Gaunama atsitiktinė įvestis;
- Generatoriaus tinklas paima atsitiktinę įvestį ir sugeneruoja naujus duomenis;
- Duomenys yra pateikiami diskriminatoriui, kuris patikrina ir suklasifikuoja sugeneruotus ir tikrus duomenis. Iš kurių grąžina rezultatą generatoriui;

- Generatorius gavęs atsakymą perskaičiuoja praradimo rezultatą, atsižvelgiant ar pavyko apgauti diskriminatorių ar generatorius bus baudžiamas už tai, kad nesugebėjo apgauti diskriminatoriaus.

Nekartą buvo minėta, kad generatorius gauna tam tikros formos atsitiktinę įvestį bet iš kur ji atsiranda? GAN įvestis naudoja atsitiktinį triukšmą (*angl. noise*). Tada generatorius paverčia šį gautą triukšmą reikšminga duomenų išvestimi. GAN gaudamas triukšmą, kaip įvestį, gali sugeneruoti įvairius duomenis, o pakeitus duotąjį triukšmą rezultatai gali būti gaunami visiškai kitokie. Nors pagal duotą skirtingą triukšmą yra sugeneruojami skirtingi vaizdai, eksperimentai rodo, kad triukšmo intensyvumo pasiskirstymas neturi didelės reikšmės rezultatuose, todėl galima pasirinkti triukšmo šaltinį iš kurio lengva paimti mėginius [GOO22].

Norint išmokyti neuroninį tinklą, reikia pakeisti tinklo svorį, kad būtų sumažintas jo išvesties klaida arba praradimas. GAN generatorius nėra tiesiogiai susijęs su apskaičiuojamais nuostoliais, kuriuos yra bandoma paveikti. Generatorius patenka į diskriminatoriaus tinklą naudojant *backpropagator*, o diskriminatorius sukuria išvestį, kurią bando paveikti. Generatoriaus yra baudžiamas ir perskaičiuojama prarastis už generatoriaus pavyzdį, kurį diskriminatoriaus tinklas klasifikuoja kaip netikrą. *Backpropagator* koreguoja kiekvieną svorį tinkama kryptimi, apskaičiuodamas svorio poveikį rezultatui – kaip pasikeis rezultatai, svoris bus pakeistas į vieną ar kitą pusę. Generatoriaus svorio poveikis priklauso nuo diskriminatoriaus svorio poveikio, į kurį jis patenka. Taigi *backpropagator* prasideda nuo išvesties ir grįžta per diskriminatorių į generatorių. Todėl kai mokosi generatoriaus neturi keistis diskriminatorius, nes keičiantis diskriminatoriui būtų dar sudėtingiau mokyti generatorių apgauti diskriminatorių. Pilna generatorius apmokymo eiga yra tokia:

- Gaunama minėta triukšmo įvestis;
- Sugeneruojama išvestis iš gauto triukšmo;
- Gauta generatoriaus išvestis pateikiama diskriminatoriui, kuris nusprendžia ar tai yra „Tikras“ ar „Netikras“;
- Apskaičiuojamas nuostolis pagal diskriminatoriaus klasifikaciją;
- Atsakas yra grąžinamas naudojantis *backpropagate* per diskriminatorių į generatorių, taip gaunant reikšmes pasikeitimams;
- Pagal gautas reikšmes keičiami generatoriaus svoriai.

Tai yra viena generatoriaus mokymo iteracija [GAM+14][GOO22][KST+22].

1.1.4 GAN nuostolio funkcija

GAN nuostolio funkcija bando nustatyti ar neuroniniai tinklai gerai mokosi. Nuostolio funkcija yra apskaičiuojama pagal 1 formulę. Gautas atstumo rezultatas yra sulyginamas su numatytu norimu rezultatu, kurio yra tikimasi. GAN modelis skaičiuoja dviejų nuostolių funkcijos rezultatus: vieną generatoriaus mokymui ir kitą diskriminatoriaus mokymui. Nuostolio funkcijoje, kuri bus apžvelgta vėliau, generatoriaus ir diskriminatoriaus nuostoliai atsiranda iš vieno atstumo tarp tikimybių skirstinių mato. Formulėje generatorius gali paveikti tik vieną dalį – netikrų duomenų pasiskirstymą, todėl generatoriaus mokymo metu atsisakoma kitos dalies, kuri atspindi realių duomenų pasiskirstymą. Galų gale generatoriaus ir diskriminatoriaus nuostoliai atrodo skirtingai, net jei jie atsiranda dėl vienos formulės. Toliau bus pateikiama nuostolio funkcija GAN modulio architektūroje, kur generatorius bando sumažinti šią funkciją, o diskriminatorius bando ją maksimaliai padidinti žr. formulė (1). GAN generatorius toliau bus vadinamas G, o diskriminatorius vadinamas D.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log(D(x))] + E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (1)$$

Šioje funkcijoje (1) matome:

- $D(x)$ yra diskriminatoriaus tikimybės įvertis, kad tikrų duomenų pavyzdys x yra tikras;
- E_x yra visų realių duomenų egzempliorių numatoma vertė;
- $G(z)$ yra generatoriaus išvestis, kai pateikiamas triukšmas z ;
- $D(G(z))$ yra diskriminatoriaus įvertis tikimybės, kad netikras egzempliorius yra tikras;
- E_z yra tikėtina visų atsitiktinių generatoriaus įvesčių vertė (tikėtina visų sugeneruotų netikrų atvejų $G(z)$ vertė).

Formulė gaunama iš kryžminės entropijos tarp tikrojo ir generuoto skirstinio. Generatorius negali tiesiogiai paveikti $\log(D(x))$ reikšmės funkcijoje, todėl generatoriaus nuostolių sumažinimas yra tolygus $\log(1 - D(G(z)))$ mažinimui, kur generatorius bando pasiekti tikimybę 1, o diskriminatorius bando didinti tikimybę, kad būtų 0. Taip kartu keisdami savo tikimybės palaiko pusiausvyrą tarpusavyje [GAM+14][GOO22].

1.1.5 GAN trūkumai ir problemos

Generatyviniai besivaržantys neuroniniai tinklai gali pasiekti išties įspūdingų rezultatų. Gali sugeneruoti naujus duomenis, atitinkančius realius objektus, sukurti tikras nuotraukas. Norint gauti kokybiškus rezultatus naudojant GAN, reikia juos tinkamai apmokyti. Tačiau ne

visados yra lengva tai atlikti, nes GAN tinklus gali būti išmokyti sunku. Viena iš pagrindinių GAN problemų ir yra, tai kad nėra iš karto aišku ar diskriminatorius ir generatorius bus gerai sinchronizuoti mokymo metu. Jei generatorius mokosi greičiau nei diskriminatorius, yra gaunamas „the Helvetic scenario“ atvejis, kurio metu nepavyksta tinkamai apmokyti generatoriaus, nes jis sugeba sugeneruoti tik mažą duomenų įvairovę. Rezultate gaunasi, kad apmokytas generatorius generuoja tik kelis ir tuos pačius rezultatus, kad ir koks įvesties triukšmas būtų paduodamas. Generatorius visada bando rasti vieną išvestį, kuri diskriminatoriui atrodo labiausiai tikėtina, jei generatorius pastoviai pradeda gaminti tą patį rezultatą (arba nedidelį rezultatų rinkinį), geriausia diskriminatoriaus strategija yra išmokti visada atmesti tą rezultatą. Bet jei naujos kartos diskriminatorius įstrigo vietiniame minimume ir neranda geriausios strategijos, tada kitai generatoriaus iteracijai yra lengva sugeneruoti įtikinamą rezultatą. Kiekviena generatoriaus iteracija per daug optimizuoja tam tikrą diskriminatorių, o diskriminatorius niekada nesugeba išėiti iš spąstų. Dėl to generatorius ir nesugeba sugeneruoti didesnio kiekio įvairovės. Ši GAN gedimo forma vadinama režimo žlugimu (*angl. Collapse*) [GAM+14][GOO22].

Tačiau gali būti ir atvirkščiai, kad generatorius negalės pagerėti, nes diskriminatorius yra stipriai pažengęs į priekį. Su dideliu tikslumu gali atmesti visus, generatoriaus pateiktus rezultatus, o generatorius negali iš jo naudingai mokytis. Tai yra vadinama nykstančio gradiento problema (*angl. vanishing gradient problem*) [GAM+14][GOO22].

Viena iš problemų, kuri nėra tokia kritinė bet gali iškilti, kai abu neuroniniai tinklai yra pasiekę savo maksimalų potencialą tai, kai generatorius, kuria visiškai realistinius duomenis ir diskriminatoriui lieka spėlioti 50/50 šansu ar pateikti duomenys yra tikri ar ne [GAM+14].

Dar vienas iš minusų GAN modelyje yra hiperparametrų (*angl. Hyperparameters*) parinkimas, kadangi jie yra labai svarbūs ir gali ilgai užtrukti išrinkti tinkamus, nes nesutvarkius teisingai hiperparametrų GAN modelis nesugeneruos gerų rezultatų [GAM+14].

Toliau bus pateikiami GAN panaudojimo būdai.

1.2. Veido generavimo panaudojimo būdai

Veido generavimas gali būti pritaikytas įvairiose situacijose pavyzdžiui norint užpildyti anketą, kurioje yra prašoma įkelti nuotrauką arba jei reikia mokslo tikslams užpildyti įsivaizduojamo žmogaus CV ir t.t. Veido generavimas padeda tokiais atvejais panaudoti žmogaus veidą nepažeidžiant kito žmogaus privatumo, kai yra naudojama jų nuotrauka, nes sugeneruotos

nuotraukos neatspindi jokio realaus žmogaus. Pavyzdžiui naudojantis puslapiu <https://thispersondoesnotexist.com/> kiekvieną kartą perkrovus puslapį yra atvaizduojamas vis kitas netikro žmogaus sugeneruotas veidas. Tačiau jei reikia kažko konkrečiau, kaip jaunesnio ar vyresnio žmogaus atvaizdo su tam tikrais veido bruožais, tam yra <https://generated.photos/>, kuris leidžia sukurti veidą pagal įvairius parametrus, kaip: amžius, emocija, plaukų ir odos spalva, lytis, galima net pridėti akinius, kepurę ir kitus aksesuarus. Taip pat „Generated Photos“ net suteikia API, kuriuo pasinaudojus galima integruoti veidų generavimą savo reikmėms. Tačiau anketų ir CV pildymas nėra vieninteliai veido generavimo panaudojimo būdai. „Generated Photos“ duomenis galima panaudoti ir kitose srityse tokiose, kaip:

- **Kūrybiniuose darbuose** – sukuriant daug ir įvairių veidų turinčių skirtingus bruožus ir aspektus;
- **Elektroninėje prekyboje** – kur svetainėje gali būti rodoma vis kitokia nuotrauka, atsižvelgiant į lankytojo profilį. Nuotraukose esantis modelis galėtų atrodyti vyresnis vyresnio amžiaus lankytojams, gali atsižvelgti į lytį ir pan.;
- **Akademiniai tyrimai** – norėdami patikrinti hipotezę, galima sukurti to paties veido variantus arba sukurti skirtingus veidus. Tyrinėdami suvokimo šališkumą, sužinoti, kaip žmonės formuoja pirmąjį įspūdį ir kt. Taip pat galima naudoti neuroninių tinklų mokymui iš atrinktų sugeneruotų rinkinių;
- **Žaidimuose** – sukuriant vis kitokių unikalų veidą žaidėjo veikėjui arba kompiuterio valdomiems veikėjams [SZS+20].

Žinoma šie naudojimo metodai yra gerieji ir nekeliantys pavojaus, tačiau sugeneruotus veidus galima panaudoti ir blogiems tikslams – tokiems, kaip netikrų žmonių anketų kūrimas, su kuriomis vėliau yra skleidžiamos melagingos naujienos, bandoma apgauti žmones ar tyčiotis iš jų.

Žmonių veidams generuoti galima naudoti tokius GAN modelius:

- StyleGAN [KLA19], StyleGAN2 [KLA+20] bei StyleGAN3 [KAL+21];
- PG-GAN [KAL+18];
- BigGan [BDS19].

1.3. Veido generavimo iš kelių nuotraukų panaudojimo būdai

Veido sujungimo iš kelių nuotraukų pagrindinė panaudojimo idėja yra tai, kad pateikus savo arba kito žmogaus veidą jį galima sujungti su dar vieno žmogaus veidu. Taip sujungiant du

veidus yra gaunamas naujas veidas turintis bendrus bruožus iš tų žmonių, kurie buvo pateikti. Internetu atlikti šiam veiksmui galima rasti įvairių puslapių, kurie teikia tokias paslaugas. Juose galima įkelti savo nuotrauką, o kitą nuotrauką galima pasirinkti iš pateikto žinomų žmonių sąrašo, taip gaunant bendrų bruožų turintį veido vaizdą (<https://www.morphthing.com/>). Be to yra puslapių, kuriuose galima gauti sugeneruotą vaiko veidą pagal pateiktus dviejų žmonių veidus (<https://baby-ac.com/>). Šios paslaugos yra teikiamos ne tik specifiniuose puslapiuose, bet ir telefono programėlėse tokiose, kaip: „MixBooth“ iOS telefonams ir „Face Morph“ android telefonams, kurie taip pat atlieka veidų sujungimą. Veidų sujungimo pavyzdys matomas 14 pav. jame yra pateikiami šaltiniai – žmonių veidai ir norimas tikslas – skirtingų žmonių veidai. Sujungus šaltinio ir tikslo veidus gaunamas bendras rezultatas su bruožais paimtais iš abiejų nuotraukų, kaip plaukų ar odos spalva, turimi aksesuarai ir t.t. Pavyzdžiui sujungus vidurinę tikslo nuotrauką su vidurine šaltinio nuotrauka, galima matyti, kad yra atvaizduojamas berniukas su bruožais, tokiais, kaip plaukų spalva, akimis, paimtais iš moters. Žinoma skirtingi algoritmai gali gauti skirtingus rezultatus, šiuo atveju rezultatai buvo gauti naudojant StyleGAN [KLA19].

Vieni GAN gali sujungti visus jungiamų nuotraukų veido bruožus, bet yra GAN, kurie leidžia pasirinkti tam tikrus veido bruožus sujungiant nuotraukas tarpusavyje. Pavyzdžiui norint sujungti vyresnio amžiaus ir jauno žmogaus veidus ir gauti jauno žmogaus veidą su vyresnio žmogaus bruožais reikia nurodyti, kokius bruožus sujungti. Arba norint pamatyti, kaip žmogus be akinių atrodys su akiniais, galima pateikti to žmogaus nuotrauką ir kito žmogaus nuotrauką, kuris turi akinius ir pasirinkus aksesuarų sujungimą, gaunamas vaizdas žmogaus su akiniais. Šio metodo pavyzdys gali būti matomas 11 pav. jame matomi sugeneruoti veidai: kairėje pusėje yra pateikiamos originalios nuotraukos, kurios toliau yra pakeičiamos pagal tam tikrus atributus, kaip:

- poza, kur nuotraukoje yra atvaizduojamas asmuo pasisukęs šonu;
- amžių, kur nuotraukoje veidas yra pasendinamas;
- emociją, kur nuotraukoje veido išraiška yra pakeičiama;
- nuotraukoje žmogui pridedami tokie aksesuarai, kaip akiniai.

Tai labiau atrodo kaip nuotraukos manipuliavimas, nei nuotraukų sujungimas, bet tai yra kelių veidų sujungimas pagal tam tikrus bruožus. Kadangi apmokant algoritmą yra analizuojamos įvairios nuotraukos pagal tam tikras kategorijas. Po apmokymų, generuojant naują veidą pagal norimą atributą yra sujungiamas pateiktas veidas ir bruožas iš kurių gaunamas sugeneruotas naujas veidas [SGT+20].

Pagrindė šią technologiją žmonės naudoja savo pramogai, pavyzdžiui pasižiūrėti, kaip jie atrodytų vyresni arba jei nešiotų akinius, ar kaip atrodytų jų ir kito žmogaus vaikas. Tačiau ne visi žmonės naudoja tokias technologijas pramogai, todėl ir Norvegijos mokslo ir technologijų universitete (*angl. Norwegian University of Science and Technology, NTNU*) kilo idėja atlikti tyrimą, kuriame buvo tikrinama ar sugeneruotos nuotraukos naudojant veido sujungimo technologijas negali būti panaudotos prieš veido atpažinimo technologijas. Taip bandant sukurti veidą, kuris tiktų dviejų pateiktų žmonių veidams atpažinti. Galutiniuose rezultatuose paaiškėjo, kad norint apeiti veidų atpažinimo saugos sistemas, gali kilti rizika, kai yra naudojama su StyleGAN sujungta veidų nuotrauka, nes sugeneruotas veidas gali tikti dviem skirtingiems asmenims, pateikiant tą patį dokumentą [SZS+20].

Be StyleGAN yra ir daugiau veidus sujungiančių generatyvinių modelių, tokių kaip:

- InterFaceGAN [SGT+20];
- BlendGAN [MQZ+21];
- ir kt.

1.4. Veido generavimo arba keitimo panaudojimo būdai vaizdo įrašė

Veido keitimas vaizdo įrašė arba kitaip dar vadinamas *Deepfake* yra paplitęs dėl savo panaudojimo galimybių. Jis turi nemažai panaudojimo būdų, kurių paskirtis gali būti tiek pramogai, mokslui, tiek dezinformacijai skleisti ar patyčioms, kas nėra priimtina, tačiau vis tiek atsiranda žmonių, kurie taip elgiasi. Šiame skyriuje bus aptariami *Deepfake* technologijos panaudojimo būdai filmų industrijoje, švietimo srityje, žaidimuose, pramogų srityje, socialinėje žiniasklaidoje, sveikatos ir mokslo srityse bei įvairiose verslo srityse, tokiose kaip mada ir elektroninė prekyba.

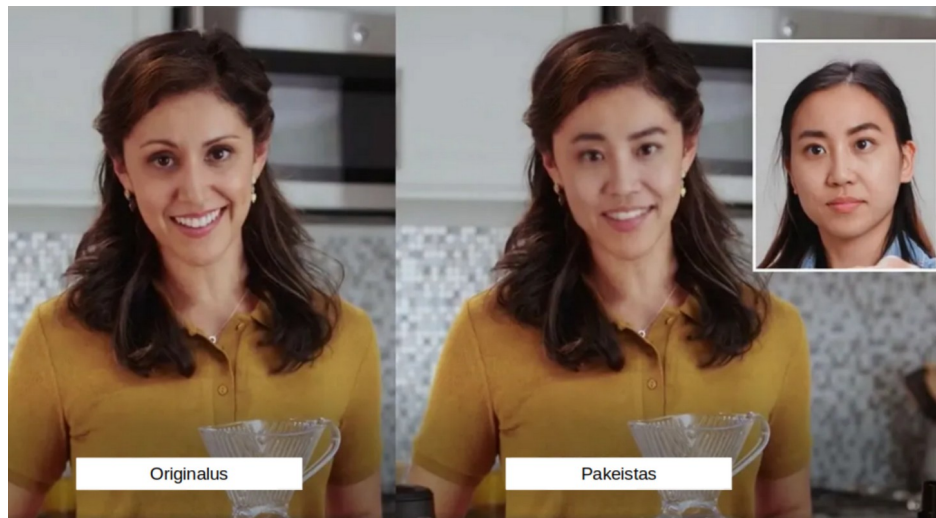
Deepfake technologija kino pramonėje gali būti panaudojama keliais būdais – pavyzdžiui, su šia technologija galima ne tik pakeisti ar sukurti veidus vaizdo įrašė, bet galima sukurti skaitmeninius balsus aktoriams, kurie neteko savo dėl ligos, arba norint atnaujinti filmuotą medžiagą, o ne filmuoti iš naujo. Kino kūrėjai gali atkurti klasikines filmų scenas, kurti naujus filmus, kuriuose vaidina seniai mirę aktoriai [YK21]. Taip pat galima panaudoti specialiuosius efektus ir veidų redagavimą atliekant montażą ir patobulinti mėgėjiškus vaizdo įrašus iki profesionalios kokybės. *Deepfake* technologija suteikia galimybę kelių žaidėjų žaidimuose ir virtualių pokalbių pasaulyje generuoti veiduose judesius atsižvelgiant į kalbančiojo atvaizdą bei pakeisti jo balsą pagal pasirinktą žaidėjo avatarą [WES19][LMH22]. Panašiu principu

technologija gali būti naudinga socialinėje bei medicinos srityse. *Deepfake* gali padėti žmonėms susitaikyti su artimųjų praradimu. Skaitmeniniu būdu prikeliant mirusį asmenį atgal į gyvenimą ir taip potencialiai padedant gedinčiam žmogui atsisveikinti. Be to, jis gali skaitmeniniu būdu sugeneruoti vaizdą ar vaizdo įrašą, kuriame gali būti vaizduojami asmenys su atkurtomis asmens galūnėmis, jei jos buvo amputuotos arba leisti žmonėms pamatyti, kaip asmuo atrodytų su kitai lyčiai būdingais bruožais. Taip pat *Deepfake* technologija gali padėti Alzheimerio liga sergantiems žmonėms bendrauti, sukuriant veidus iš tų laikų, kai sergantis žmogus dar prisimena tam tikrą asmenį. Mokslininkai taip pat tiria GAN panaudojimą rentgeno spindulių anomalijai aptikti ir jų potencialą kuriant virtualias chemines molekules, kad pagreitintų medžiagų mokslą ir medicinos atradimus.

Įmonės domisi prekės ženklui taikomų *Deepfake* technologijų potencialu, nes jos gali reikšmingai pakeisti el. prekybą ir reklamą. Pavyzdžiui, prekės ženklai gali sudaryti sutartis su įmonėmis, kurios siūlo sukurti virtualius modelius, kurie galėtų reklamuoti madingą aprangą ar kitokį prekės ženklą su įvairiausiomis virtualaus modelio kombinacijomis, t. y. Virtualūs modeliai gali turėti skirtingą sudėjimą, odos atspalvį, ūgį ar svorį. Be to, *Deepfake* leidžia pačiam patapti modeliu – pagal savo nuotrauką *Deepfake* gali sugeneruoti atvaizdą, kaip asmuo esantis nuotraukoje atrodytų būdamas modeliu. Sukūrus savo modelį yra atrakinama dar daugiau *Deepfake* galimybių. Pavyzdžiui galimybė peržiūrėti aprangą prieš ją perkant ant savo modelio, kad būtų įmanoma įsivaizduoti ir pamatyti, kaip ši apranga atrodytų realybėje ant realaus žmogaus. Taip pat *Deepfake* gali generuoti tikslinius mados skelbimus, kurie skiriasi priklausomai nuo laiko, oro ir asmens, kuris žiūri skelbimus. Be to, ši technologija gali pateikti unikalius dirbtinius balsus, kurie išskiria įmonės produktus video reklamoje, kad būtų lengviau atskirti prekės ženklą, tokį funkcionalumą taip pat galima panaudoti filmų industrijoje [WES19].

Kaip ir buvo minėta šią technologiją galima pritaikyti naudingai įvairiose srityse, nepaisant to *Deepfake* technologija gali kelti didesnę pavojų negu prieš tai aptarti veidų generavimo metodai. Naudojant šią technologiją galima sukurti melagingus vaizdo įrašus, kuriuose be žmogaus sutikimo specifinis žmogus atlieka veiksmus ar kalba, tai kas buvo nurodyta generuojant vaizdo įrašą. Dažniau naudojamas ir populiariesnis būdas yra ne generuoti pilną įrašą nuo pradžių iki galo, o jau esamame vaizdo įrašė pakeisti tikslinio žmogaus veidą kitu, norimu, žmogaus veidu [ETI21][WES19]. Toks manipuliavimas žmonių atvaizdais be jokio sutikimo sukėlė ne vieną socialinių tinklų ir žiniasklaidos susirūpinimą, nes vis dažniau pasitaiko melagingų vaizdo įrašų, kuriuose vaizduojami politikai ar prezidentai, kurių kalbos buvo

iškraipytos, pakeistos dezinformacijos tikslais. Pavyzdžiui, buvo paskelbtas vaizdo įrašas socialiniuose tinkluose, kuriame buvęs Amerikos prezidentas Barakas Obama kalba apie *Deepfake* galimybes sukurti vaizdo įrašus, kur bet koks žmogus gali sakyti bet ką, ką panorėjo kūrėjas. O pats vaizdo įrašas ir buvo sukurtas naudojant šia technologija ([nuoroda į vaizdo įrašą](#)). Toliau pateikiamas pavyzdys, kaip atrodo *Deepfake* vaizdo įrašo kadras 5 pav.



5 pav. Veido pakeitimas vaizdo įrašė. Šaltinis: spectrum.ieee.org

Paveiksle matomas vaizdo įrašo kadras, kuriame vaizduojamas originalus kadras kairėje ir dešinėje pakeistas veidas. Taip pat iš paveikslo matoma, kad naujasis veidas yra šiek tiek neryškus, tačiau vis tiek įtikinamas.

Veido pakeitimui galima rasti nemažai tokių GAN modelių, kaip:

- Video2StyleGAN [AZM+22];
- FSGAN [NKH19];
- ReenactGAN [WZL+18];
- UniFaceGANs [CHW+21];
- DeepFaceLab [PGC+21].

1.5. Taikymo sričių ir originalaus GAN apibendrinimas

Buvo aptartas originalus GAN, kurio konceptas buvo pateiktas 2014 metais, kad generatyviniai besivaržantys neuroniniai tinklai suteikia naują būdą neuroniniams tinklams apmokyti. Šis būdas yra įvardinamas, kaip dviejų žaidėjų žaidimas, kuriame vienas žaidėjas kuria naują informaciją bandant atkartoti realią informaciją (generatorius), o kitas žaidėjas bando sukurtą informaciją atrasti ir klasifikuoti kaip netikrą (diskriminatorius). Tačiau gali būti

susiduriama su mokymo problemomis, kurios neigiamą įtaką galutiniams rezultatams. Generatorius gali pralenkti mokyme diskriminatorių ir atvirkščiai. Taip atsitikus rezultate yra gaunami mažos įvairovės arba netikslingai generuojami duomenys. Tačiau kaip ir buvo paminėta jau yra nemažai sukurtų GAN modelių, kurie bando išspręsti šias problemas ir kitaip patobulinti GAN modelį, kad būtų gaunami geresni rezultatai.

Taip pat buvo apžvelgta veido generavimo, panaudojamo būdai ir sritys kaip mokslo, medicinos, filmų industrijos, žaidimų, pramogų ir elektroninių parduotuvių. Kiekvienas veido generavimas, keitimas vaizde ar vaizdo įrašė turi savo panaudojimą. Veido generavimas gali padėti suasmeninti el. parduotuvėje esančius produktų atvaizdavimus ar sukurti žaidimo charakterių veidus. Kelių veidų sujungimas gali parodyti, kaip atrodys dviejų žmonių vaikas arba leisti pamatyti, kaip vienas ar kitas žmogus atrodys senatvėje arba su tam tikrais aksesuarais. Bei veidų pakeitimas vaizdo įrašė, gali padėti filmų industrijoje atgaivinti seniai mirusius aktorius naujuose filmuose. Tačiau kiekvienas turi ir savo keliamų pavojų, kaip melagingos informacijos skleidimas arba kitų žmonių teisių pažeidimas naudojant jų veidus be sutikimo interneto erdvėje.

Norint geriau suprasti, kaip veidų generavimas, jo sujungimas ar pakeitimas vaizdo įrašė veikia, reikia juos detaliau išanalizuoti, todėl toliau yra pateikiami GAN, kurie bus nagrinėjami, jų veikimas bei architektūriniai pakeitimai lyginant su originaliu GAN:

- Veidų generavimui:
 - StyleGAN2 pasirinktas, nes yra vienas iš dažniau pasitaikančių GAN veidų generavime atsižvelgiant į palyginimus straipsniuose, jo paminėjimą ar naudojimą tam tikrose paslaugose. [KLA+20];
 - PG-GAN buvo pasirinktas, nes jis taip pat dažnai randamas straipsniuose bei dėl šio modelio papildomai pasiūlyto būdo, kaip padidinti ryškumą nuotraukose didinant sluoksnių skaičių mokymosi metu, padaryti mokymosi metodą stabilesniu bei padidinti nuotraukų variaciją [KAL+18];
 - BigGan buvo pasirinktas, nes pateikia sprendimą, kaip sugeneruoti didesnės rezoliucijos nuotraukas bei kaip jas padaryti realistiškesnes pridėjus tam tikrus parametrus ir padidinant mastelį mokymosi modelyje [BDS19].
- Veidų sujungimui:
 - StyleGAN buvo pasirinktas dėl tokių pačių priežasčių kaip ir StyleGAN2 [KLA19];
 - InterFaceGAN pasirinktas, nes siūlo modelį, kuris atpažįsta tam tikrus veido bruožus kaip amžių, plaukus, aksesuarus. Naudojant jį galima sugeneruoti veidą su norimais

atributais netgi naudojant vaizdus, kurie buvo sugeneruoti kitų GAN modelių tokių, kaip StyleGAN arba PG-GAN [SGT+20];

- BlendGAN pasirinktas, nes siūlo naują modelį, kuris gali savarankiškai pritaikyti norimam veidui artistinių stilių pagal pateiktą pieštą, tapytą ar bespalvį portretą. Iš portreto yra paimamas artistinis stilius ir pritaikomas veidui. Kituose GAN modeliuose, jei norima pritaikyti tam tikrą piešinio stilių, reikia juos mokinti ties kiekvienu norimu stiliumi [MQZ+21].
- Veidų pakeitimui vaizdo įrašė:
 - FSGAN, pateikia būdą generuoti vaizdo įrašus pakeičiant veidus realiu laiku ir tai galima atlikti neapmokant GAN modelio kiekvieną kartą iš naujo. Taip pat šis GAN modelis generuoja aukštos kokybės rezultatą [NKH19];
 - DeepFaceLab, šis GAN buvo pasirinktas, nes jis vis dar yra plėtojamas, kaip atviro kodo produktas [PGC+21].

2. Generatyvinių besivaržančių neuroninių tinklų analizė

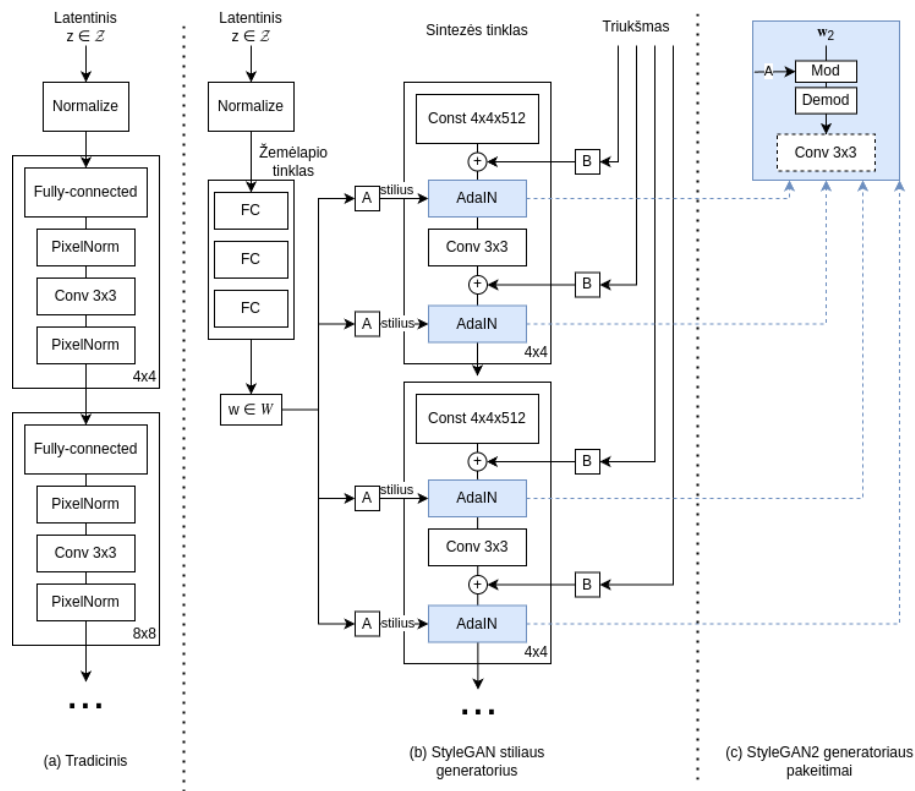
Kaip ir buvo minėta, šiame skyriuje bus analizuojami StyleGAN2, PG-GAN, BigGAN modeliai veidų generavime. StyleGAN, InterfaceGAN, BlendGAN modeliai veidų sujungime. DeepFaceLab, FSGAN modeliai veidų sujungime vaizdo įrašuose. Bus apžvelgti GAN modelių architektūros ir veikimas.

2.1. Veido nuotraukas generuojančių GAN analizė

Šiame poskyryje bus analizuojami ir lyginami GAN, kurie generuoja veidus, t.y. StyleGAN2, PG-GAN ir BigGAN. Taip pat bus apžvelgiama, kaip kiekvienas iš nurodytų GAN pakeičia originalaus GAN veikimą.

2.1.1 StyleGAN2 analizė

Iš pradžių bus aptariamas StyleGAN – kas buvo pakeista lyginant su originaliu GAN modeliu, kad suprasti, kodėl šiai kategorijai buvo pasirinktas StyleGAN2. Taigi StyleGAN modelis, įtraukia į nuotraukų generavimą nemažai naujų idėjų ir pasiūlymų. Visos naujovės ir idėjos yra GAN generatoriaus, o ne diskriminatoriaus, dalyje. Diskriminatoriaus dalis yra palikta originali. Generatoriaus dalyje atlikti pakeitimai leidžia sugeneruoti norimo stiliaus veidą su specifiniais bruožais, ko su paprastu GAN modeliu padaryti tiesiog neišėdavo. Paprastas GAN modelis kiekvieną kartą generuodamas veidą gali gauti vis kitokį rezultatą, kuris priklauso nuo pradinio įvesties parametro, tačiau galimybės pasirinkti bruožus nėra. Kitaip tariant tokiu generavimo būdu nėra jokios galimybės valdyti generuojamo rezultato bruožų, jei norima sukurti veidą sakykime moters su rudais plaukais ar vyresnio vyriškio. Todėl StyleGAN pateikia būdą, kaip būtų galima valdyti žmogaus veido stilių, kuriuo naudojantis būtų galima valdyti gaunamų rezultatų bruožus. Rezultatus galima valdyti pasirinkus tam tikrus parametrus, tokius kaip: poza, veido strazdanos, plaukai, odos spalva, lytis ir t.t., prieš pat generavimą. Generavimo metu atsižvelgiant į nustatytus parametrus yra sugeneruojamas unikalus veidas su nustatytu stiliumi. Kaip tai atrodo generavimo schema galima matyti 6 pav. [KLA19].



6 pav. StyleGAN generatoriaus architektūra [KLA19]

Šiame paveiksle yra pavaizduota tradicinio GAN ir StyleGAN generatorių architektūra. Tradicinio generatoriaus architektūroje yra paduodamas pradinis parametras, iš kurio toliau yra sugeneruojama nuotrauka. StyleGAN architektūroje galima matyti, kad yra pridėtos 3 naujas dalys: pirmoji dalis yra žemėlapio tinklas, kuriame pateikiami keli pradiniai vektoriai, kurie yra priskiriami naujiems vektoriams w . Šie vektoriai toliau panaudojami atsitiktinai pateikiant į naują normalizacijos bloką *AdaIN*, kuris yra antra nauja dalis. Trečia nauja dalis yra atsitiktinio triukšmo (*angl. noise*) implementacija ne tik į pradžią, kas reiškia jog atsitiktinis triukšmas yra pateikimas ne tik pradžioje bet ir kiekviename kitame žingsnyje su skirtinga reikšme, taip leidžiant valdyti generuojamos nuotraukos/veido „stilių“. Toliau StyleGAN2 pakeitimas architektūroje pagrinde buvo pakeista *AdaIN*, funkcija. Galima matyti, kad ši vieta išsiskaidė į dvi dalis. Vienoje dalyje vektorius yra normalizuojamas ir pateikiamas A stilius, o kitoje paskirstomi svoriai pagal stiliaus nustatymus ir toliau pateikiami į *Conv 3x3*, kaip ir StyleGAN architektūroje.

Taigi su tokiais generatoriaus pakeitimais yra galima sukurti įvairių stilizuojamų pagal poreikį veidų. Tačiau ši architektūra nėra ideali, nes sugeneruotos nuotraukos, gali turėti tam

tikrus artefaktus, kurie sugadina nuotraukų realistiškumą bei pačias nuotraukas. Todėl yra pateikiamas atnaujintas StyleGAN2, kuriame yra pakeista architektūra, taip, kad išgautų didesnės raiškos ir mažiau sugadintų nuotraukų. Kaip ir buvo parodyta tai atliekama pakeičiant *AdaIN* normalizacijos bloką ir svorių paskirstymą, pagal nustatytus stilių parametrus [KLA+20].

Taigi StyleGAN suteikia galimybę sugeneruoti realistiškus veidus su tam tikrais stiliais iš pateiktų parametrų, tačiau ne visos nuotraukos yra labai aukštos kokybės. Kai kurios sugeneruotos nuotraukos gali turėti artefaktų ar kitų netinkančių detalių, todėl yra pateikiamas naujas StyleGAN2. Šis modelis sprendžia aptartas problemas taip, kad būtų sugeneruojama kuo mažiau nuotraukų su artefaktais.

2.1.2 BigGAN analizė

BigGAN modelį aprašančiame straipsnyje [BDS19] pateikiama, kad bendras GAN modelis turi didžiulę naudą didinant nuotraukos mastelį. Mokymo dalyje pakeliant dvejetainis ar net keturiais kartais parametrus ir padidinant aštuonis kartus partijos dydį (*angl. Batch size*), gaunami geresnės kokybės rezultatai. Taip pat BigGAN pakeičia architektūrą padidinant mastelį, pagerinant kondicionavimą ir padidinant našumą. Su šiais pakeitimais BigGAN modelis gali sugeneruoti realistiškus vaizdus, veidus ir kitus objektus, nes bendras modelis buvo mokomas su dideliu kiekiu nuotraukų, kuriose buvo daugiau nei 1000 kategorijų. Nors BigGAN buvo treniruojamas su daug duomenų šis modelis turi trūkumų: pačio modelio apmokymas reikalauja labai daug resursų ir pradinis modelis buvo mokomas su daugiau nei šimtu vaizdo plokščių, kas reiškia, jog norint apmokinti tokį modelį su asmeniniais resursais minimaliai reikia turėti 4 GPU ([GitHub](#)). Pradinis apmokytas modelis po to buvo patalpintas atvirai prieigai. Taip pat dar viena iš BigGAN problemų yra tai, kad generuojamose nuotraukose gali pasitaikyti nemažai sugadintų, turinčių artefaktų ar kitokių neatitikimų. Tai galima matyti 7 pav.



7 pav. BigGAN sugeneruotos nuotraukos [Tensorflow.org](https://www.tensorflow.org)

Paveiksle matomos sugeneruotos BigGAN modelio nuotraukos, kuriose yra atvaizduojami šunys ir vištos. Vienos nuotraukos yra realistiškos ir įtikinamos, o kitos yra iš darytos ir netikroviškos.

Taigi BigGAN turi labai didelį apmokytą modelį su daugybe kategorijų ir galimybę sugeneruoti aukštos kokybės ir itin realistiškas nuotraukas, tačiau kol sugeneruojama norima nuotrauka gali pasitaikyti nemažai sugadintų nuotraukų.

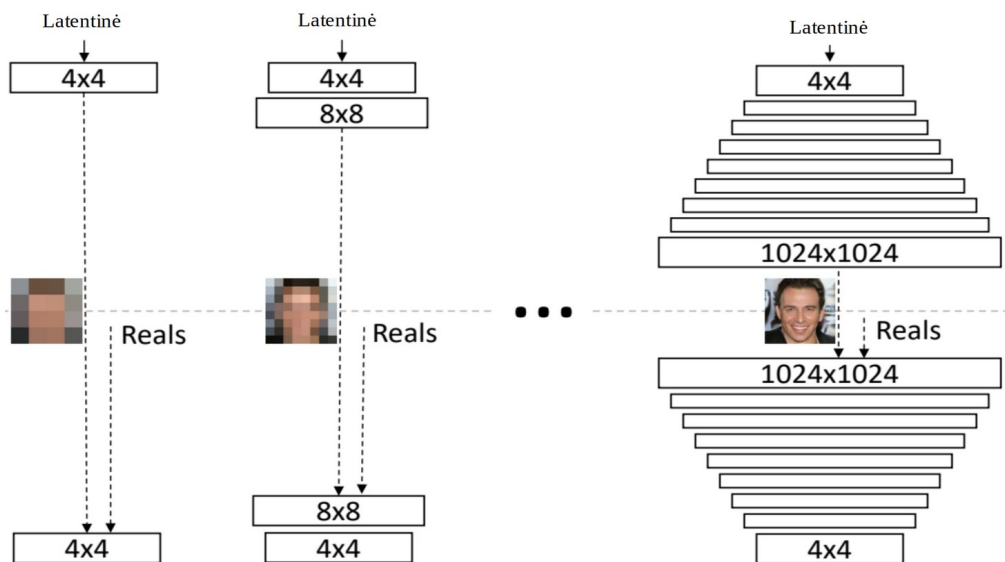
2.1.3 PG-GAN (ProGan) analizė

Progresyviai augantis GAN (*angl. Progressive Growing GAN*) arba kitaip žinomas, kaip PG-GAN ar ProGAN, keitė patį mokymo proceso atlikimą, kad jis būtų stabilesnis. Turint stabilesnį mokymo procesą yra gaunamos aukštesnės kokybės sugeneruotos nuotraukos. PG-GAN stabilesnį mokymą pasiekė pradedant nuo labai mažos rezoliucijos, tokios kaip 4x4 ir tada palaipsniui pridėdant sluoksnių blokus, kad generatoriaus modelis išvestų vis didesnės raiškos nuotrauką ir padidintų diskriminatoriaus modelio įvesties nuotraukos dydį. Toks procesas yra atliekamas tol, kol yra gaunamas norimas vaizdo dydis. Šis metodas pasirodė esąs labai veiksmingas kuriant aukštos kokybės sintetinius vaizdus, kurie yra realistiški.

Iš esmės PG-GAN pridėjo 3 pagrindinius veiksmus:

1. GAN apmokyje pridėjo progresyvų augimą (rezoliucijos didinimas);
2. Pakeitė normalizavimo funkciją į *PixelNorm*;
3. Prie diskriminatoriaus pridėjo papildomą *Minibatch* funkcionalumą;

Pirmas punktas, kaip ir buvo minėta, yra mokymo atžvilgiu pakeisti žingsniai – GAN mokymas pradedamas nuo mažos rezoliucijos pvz. 4x4 ir ties kiekviena išvestimi rezoliucija yra didinama iki norimo dydžio pvz. 1024x1024. Šio proceso diagramą galima matyti 8 pav.



8 pav. PG-GAN mokymo schema [KAL+18]

Nuotraukoje matomas progresyvus mokymo procesas. Generatorius (G) ir diskriminatorius (D) pradeda veikimą su nedidele rezoliucija. Mokymo eigoje D ir G kartu didina pačios nuotraukos rezoliuciją. Šis procesas, anot autorių, mokymo procesą gali sutrumpinti iki 2-6 kartų atsižvelgiant į tai, kokios rezoliucijos rezultato yra tikimasi. Bei naudojant tokį modelio apmokymą yra išgaunamos geresnės kokybės nuotraukos [KAL+18].

Sekantis pakeitimas yra *PixelNorm* normalizacija mokymo metu, kad generatorius ir diskriminatoriaus mokymo „lygiai“ nesiskirtų, nes kitu atveju generatorius arba diskriminatoriaus gali išmokti daugiau nei kitas ir dominuoti, taip stabdant mokymą ir gaunant netinkamus rezultatus [KAL+18].

Dar vienas pakeitimas yra pritaikytas diskriminatoriui – pridėdant autorių vadinamą *Minibatch*, kuris leidžia po kiekvienos iteracijos prisiminti prieš tai sugeneruotas nuotraukas, taip padedant diskriminatoriui nusakyti ar generatoriaus nuotrauka yra tikra ar ne. Tai priverčia generatorių generuoti įvairesnes nuotraukas, norint apgauti diskriminatorių. Šis metodas padeda išspręsti vieną iš GAN turimų problemų – apmokyti modeliai turi mažai variacijos nuotraukose.

Taigi PG-GAN pateikė pakeitimus, kurie padeda mokymą padaryti stabilesnį pridėdant normalizacijos bloką *PixelNorm*, kad nei generatoriaus ar diskriminatoriaus neišveržtų į priekį taip sugadinant mokymą. Taip pat tolygesnį mokymą padėjo palaikyti palaipsninį rezoliucijos didinimą iki nustatytos, taip padedant abiem modeliams progresyviai mokytis. Bei diskriminatoriui pridėjus *Minibatch*, buvo galima išgauti daugiau ir įvairesnių rezultatų. Visi šie pakeitimai leidžia sugeneruoti realistiškas ir aukštos kokybės nuotraukas su didesne įvairove nei

įprastai. Tačiau, mokant GAN modelį reikia atsižvelgti į nuotraukų kokybę bei kartais net į poziciją objektų poziciją, nes šio modelio mokymo metu yra „jautrus“ į tokius pokyčius, jei vieną objektai atvaizduojami yra tiesiai, o kiti šonu, gali galutiniame rezultate sugadinti generuojamus rezultatus.

2.1.4 Apibendrinimas

Šiame poskyryje buvo aptarti GAN modeliai, kurie pasiūlė visiškai kitokius sprendimus nuotraukų generavime. StyleGAN pateikia galimybę sugeneruoti realistiškus veidus ir valdyti jų stilius su tam tikrais parametrais. BigGAN buvo apmokytas su dideliu kiekiu nuotraukų ir daugybe kategorijų. Šis GAN padidino bendruosius mokymo parametrus, nuotraukų rezoliuciją ir taip išgavo aukštos kokybės realistiškas nuotraukas. Tačiau šių BigGAN pakeitimų minusas yra tai, kad mokymas reikalauja labai daug resursų. PG-GAN pateikia naują mokymo būdą, kuriame mokymas yra pradedamas nuo mažų rezoliucijos ir taip ją didinant iki norimos. Taip pat pateikė naują normalizavimo funkciją, kuri suteikia stabilumo modelio apmokymui bei diskriminatoriuje pridėjo *Minibatch*, kuris priverčia generatorių sugeneruoti įvairesnių nuotraukų.

2.2. GAN skirtų veidų generavimui iš kelių nuotraukų analizė

Šiame poskyryje bus analizuojami tokie GAN, kaip InterFaceGAN, BlendGAN bei StyleGAN. Šie GAN generuoja veidus sujungdami kelias nuotraukas.

2.2.1 InterFaceGAN analizė

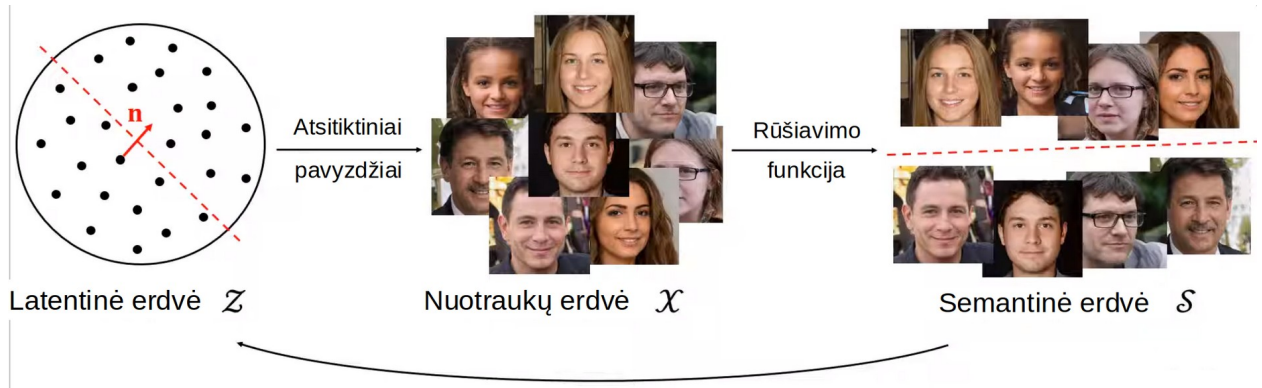
InterFaceGAN pateikia kitokį sprendimą, kaip galima manipuluoti veidus nuotraukose, nekeičiant pačio GAN generavimo ar mokymo dalių. InterFaceGAN modelio esmė yra suprasti vadinamą latentinę erdvę (*angl. Latent space*), kurioje yra saugomi GAN modelio išmokti rezultatai. Vėliau rezultatai yra suskirstomi į grupes ir naudojantis nustatytomis grupėmis toliau manipuluojami norimi veido bruožai. Taip gaunant veido manipuliavimą nemokinant GAN modelio iš naujo, jei norima atlikti skirtingus veido pakeitimus. Šio veikimo principo žingsniai yra tokie:

- Iš pradžių yra atrenkamos atsitiktinės nuotraukos iš išmokyto modelio latentinės erdvės taškų, kurioje kiekvienas taškas yra tam tikra nuotrauka;
- Toliau išrinktos nuotraukos yra sugrupuojamos į tam tikras semantines grupes, naudojant rūšiavimo funkciją. Pavyzdžiui viena grupė yra moterys, kita grupė yra vyrai ir prie

kiekvienos iš šių grupių yra priskiriamos atitinkamos semantinės žymos, nurodančios, kuriai grupei išrinktos nuotraukos priklauso;

- Turint semantines grupes, latentinėje erdvėje yra nustatoma riba tarp numatytų grupių;
- Galiausiai gaunamas semantinis veido keitimas keliant pateiktą veidą iš vienos grupės į kitą;

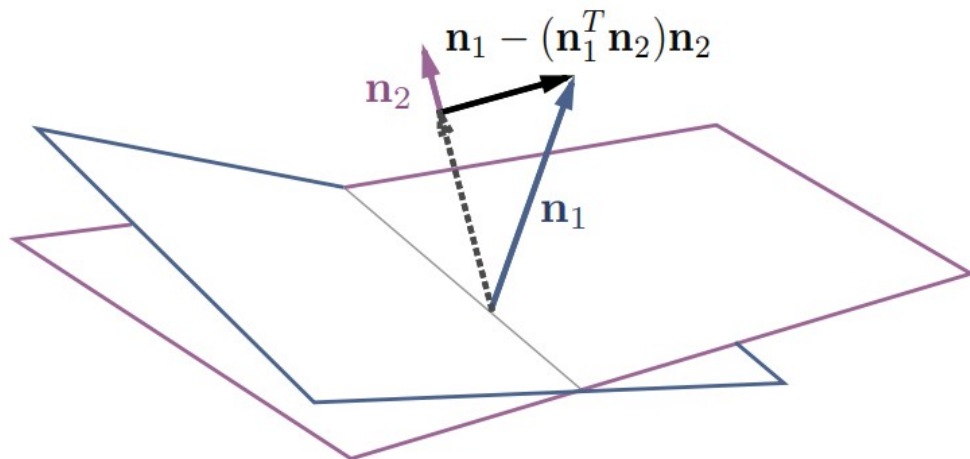
Bendras šių žingsnių atvaizdavimas gali būti matomas 9 pav. [SGT+20].



9 pav. InterFaceGAN bendras nuotraukos manipuliavimo pavyzdys [SGT+20]

Paveiksle matoma latentinės erdvės dalis iš kurios, kaip buvo minėta, atrenkamos atsitiktinės nuotraukos, kurios vėliau panaudojant rūšiavimo funkciją yra sugrupuojamos į semantines grupes. Semantinėms grupėms yra priskiriamos žymės, pagal kurias latentinės erdvės grupės yra pažymimos, taip gaunant ribas tarp tam tikrų buožių, kaip poza, lytis, amžius, aksesuarai ir t.t. Turint apibrėžtas ribas, toliau pateikus nuotrauką ji būtų priskiriama prie nustatytos grupės ir keliant pateiktą nuotrauką iš vienos grupės į kitą būtų gaunamas semantinis veido keitimas.

Taip pat buvo pasiūlytas būdas, kaip galima manipuluoti norimomis veido bruožo dalimis nekeičiant visų veido bruožų. Tai galima atlikti naudojant sąlyginį manipuliavimą (*angl. Conditional manipulation*). Taikant šį metodą galima nustatyti, kokį veido bruožą norima pakeisti, sakykime šiuo atveju norima pakeisti amžių. Jei yra naudojamas vyresnio vyro ir jaunesnės moters nuotraukos, norint pakeisti amžių yra apskaičiuojama nauja pozicija atsižvelgiant į keičiamą bruožą latentinėje erdvėje. Pagal apskaičiuotą rezultatą toliau yra keičiama pateiktos nuotraukos pozicija latentinėje erdvėje, kurios metu bus pakeistas tik norimas bruožas, o visi kiti bruožai paliekami lygiai tokie patys, kokie buvo atvaizduojami pradinėje nuotraukoje. Sąlyginis manipuliavimas suteikia galimybę lanksčiau valdyti veidų manipuliavimą ir jo schema gali būti matoma 10 pav.

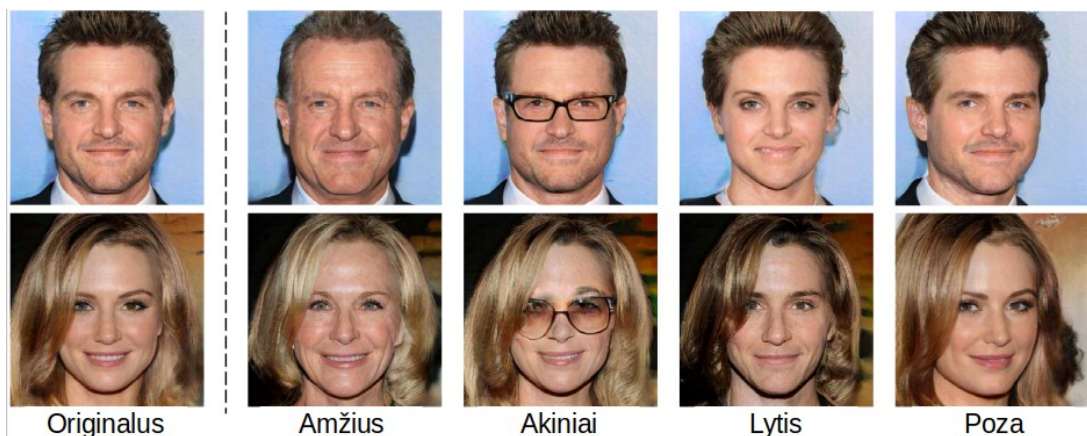


10 pav. InterFaceGAN sąlyginio manipuliavimo pavyzdys [SGT+20]

Paveiksle yra matomas sąlyginio manipuliavimo pavyzdys, kuriame yra du pagrindiniai vektoriai – vektorius n_1 ir vektorius n_2 . Vektorius n_1 nurodo pagrindinę poziciją arba kitaip, kokį veido bruožą reikia keisti. Vektorius n_2 nurodo, kaip numatytas veido bruožas turi būti keičiamas. Taip pat galima matyti bendrąją formulę, kuri yra apskaičiuojama iš nurodytų dviejų vektorių, naują kryptį norint pakeisti tik tą veido bruožą, nekeičiant kitų. Skirtingų spalvų lapai 10 pav. vaizduoja erdvėje esančias nuotraukas.

Taigi naudojant latentinę erdvę GAN modelyje, su pateiktais InterFaceGAN pakeitimais, galima keisti veido bruožus nuotraukose ištraukiant tam tikras detales arba pilnai keičiant visus veido bruožus, kurie gali būti paimti iš kitos nuotraukos. Šie metodai leidžia pakeisti pozą, amžių, lytį, uždėti akinius bei įvairiai keisti kitokius bruožus. Taip pat naudojant InterFaceGAN galima panaikinti arba pataisyti, kitų GAN sugeneruotus veidus, jei juose yra artefaktų ar kitokių problemų [SGT+20].

InterFaceGAN rezultatų pavyzdys pateikiamas 11 pav.



11 pav. InterFaceGAN rezultatų pavyzdžiai [SGT+20]

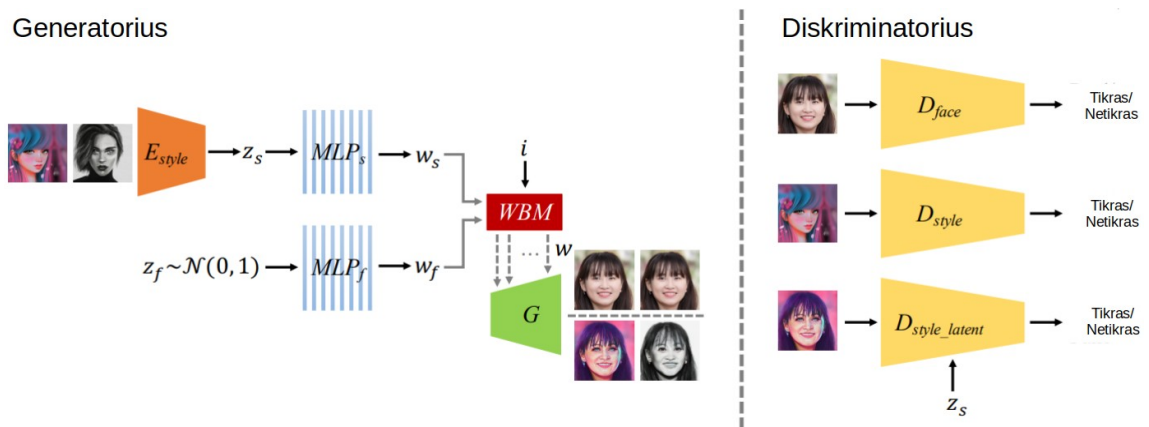
Paveiksle atvaizduoti InterFaceGAN sugeneruoti rezultatai pagal tam tikrus bruožus: amžių, pridėdamus aksesuarus, lytį bei pozą. Taip pat matoma ir originali nuotrauka, kuri buvo pakeista.

Aptariant galima teigti, jog InterFaceGAN pateikia naują būdą generuoti veidus sujungiant kelias nuotraukas. Tai atlieka naudojant latentinę erdvę iš išmoktų modelio nuotraukų, kurioje nuotraukos yra suskaidomos į tam tikras grupes ir atspindi tam tikrus veido bruožus, tokius kaip: amžius, aksesuarai, lytis, poza. Turint šias grupes yra gražinami grupių rezultatai į latentinę erdvę, kurioje sudedamos ribos ties grupėmis ir keičiant nuotraukos poziciją erdvėje keičiasi ir nuotraukos bruožai. Taip pat InterFaceGAN pateikia būdą valdyti bruožų pasikeitimą. Kad būtų keičiami ne visi veido bruožai, o tik tam tikri atributai, galima taikyti sąlyginį manipuliavimą, kurio metu paskaičiuojamas naujas taškas latentinėje erdvėje ir pagal tai pakeičiamas tik vienas veido bruožas.

2.2.2 BlendGAN analizė

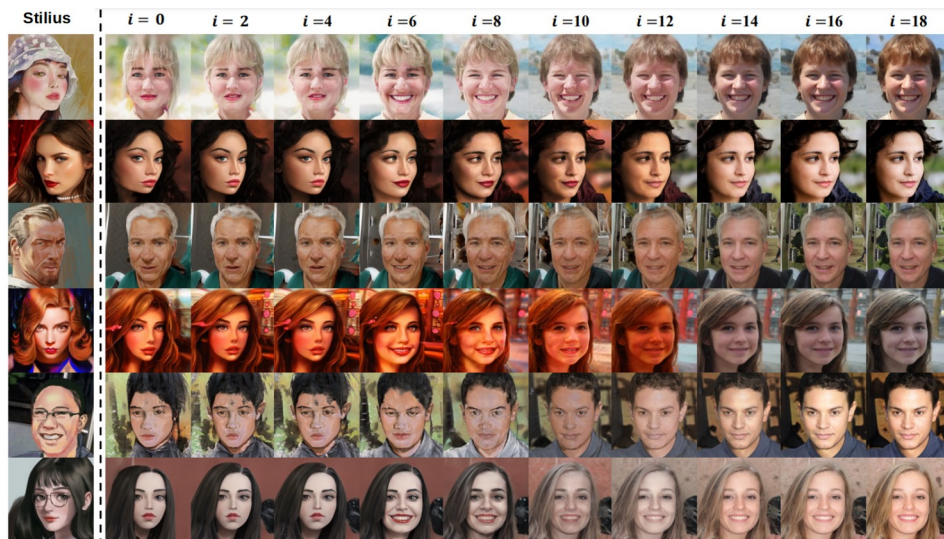
BlendGAN pateikia veidų manipuliavimo ir sujungimo būdą, kuris gali manipuluoti veido nuotraukomis pritaikydamas tam tikrą artistinį stilių. Ir nesvarbu ar tai būtų pieštuku nupieštas portretas ar tai būtų animacinis veidas. BlendGAN leidžia sujungti veido nuotrauką su pasirinktu artistiniu veido stiliumi ir rezultate gaunant unikaliai atrodantį savo ar kito žmogaus veidą su pritaikytu stiliumi. Ir ši galimybė BlendGAN išskiria iš kitų generatyvinių besivaržančių neuroninių tinklų. Norint pritaikyti skirtingus artistinius stilius nereikia iš naujo apmokyti GAN modelio, kad šis veiktų. Tokių rezultatų autoriai sugebėjo pasiekti pritaikius kelis papildomus žingsnius. Pirmiausia autoriai mokino save prižiūrintį (*angl. Self-supervised*) enkoderį (*angl. encoder*), kuris išgauna artistinį stilių iš nuotraukos. Toliau yra inicijuojamas BlendGAN pateiktas naujas modelis, kuris vadinasi *WBM* (*angl. Weighted blending module*). Šis modelis yra atsakingas už veidų sujungimą tarp stilizuoto veido ir pasirinkto veido. Jis tai atlieka sujungdamas gautus latentinius stiliaus ir veido kodus, gautas rezultatas yra atiduodamas į generatorių. Žinant veido ir stiliaus latentinius kodus, galima valdyti galutinį variantą, intensyviau ar mažiau intensyviai pritaikyti pasirinktą stilių. Taip pat *WBM* dalyje keičiant indikatorius reikšmę galima valdyti kokios veido ir stiliaus dalys bus sujungtos galutiniame rezultate. Be to BlendGAN modelis yra apmokytas naudojant *AAHQ* duomenų rinkinį, kuriame yra įvairių stilių portretų su spalvų tonais ir veidų atributais [MQZ+21].

BlendGAN veikimo žingsnių schema pateikiama 12 pav.



12 pav. BlendGAN veikimo pavyzdys [MQZ+21]

Nuotraukoje pateikiamas BlendGAN veikimo principas su generatoriumi ir diskriminatoriais. Generatoriaus pusėje yra E_{style} enkoderis, kuris išgauna stilių iš nuotraukos ir jį pateikia kaip z_s . z_f dalis priklauso veidui, kurio stilių reikia pakeisti. Ši reikšmė išgaunama naudojant „Gaussian distribution“. Toliau z_s ir z_f reikšmės yra perduodamos į daugiasluoksnį perceptroną (MLP bloką) iš kurio išgaunamos reikšmės w_s ir w_f . Šios reikšmės toliau pateikiamos į WBM modelį, kuriame yra sujungiamos veido ir stiliaus detalės. Gautas rezultatas yra pateikiamas į generatorių, iš kurio grąžinamas sugeneruotas veidas. Kaip galima matyti 12 pav. BlendGAN naudoja ne vieną, o tris diskriminatorius, kurie atlieka skirtingus tikrinimus. Pirmasis tikrina ar veidas yra tikras, antrasis diskriminatorius tikrina ar nuotraukoje yra pateiktas stilizuotas veidas ir paskutinis tikrina ar sugeneruotas stilizuotas veidas yra tokio pat stiliaus kokio buvo norima. Taip pat nuotraukoje galima matyti kintamąjį i , kuris gali būti paduodamas į WBM bloką. i yra indikatorius, kurio reikšmės gali būti nuo 0 iki 18. Ši reikšmė nurodo stiliaus pritaikymo intensyvumą. 0 reiškia pilnai stilizuotą nuotrauką, o nustačius i reikšmę 18, rezultato nuotrauka gaunama be stiliaus. Generuojami rezultatai naudojant BlendGAN pateikti 13 pav. [MQZ+21].



13 pav. BlendGAN rezultatų pavyzdžiai [MQZ+21]

Nuotraukoje galima matyti, sugeneruotus rezultatus naudojant BlendGAN modelį su skirtingomis indikatorius reikšmėmis. Kaip ir buvo minėta indikatorius reikšmės gali būti nuo 0 iki 18. Žinoma, galima pasirinkti ir didesnę indikatorius reikšmę, tačiau pasirinkus 18 jau yra gražinama nuotrauka be stiliaus, vadinasi didesnė indikatorius reikšmė nesudarys jokio žymaus skirtumo. Nustačius indikatorius reikšmę lygią 0, generuojamos nuotraukos pakeičia ne tik stilių bet ir įvairius veido bruožus atsižvelgiant į nurodytą stilių. Pasirinkus indikatorius reikšmę lygią 6, generuojamose nuotraukose yra išlaikomi originalaus veido bruožai ir pritaikomi stilius pokyčiai.

BlendGAN pateikia unikalų būdą, kaip galima sugeneruoti veidus pagal pasirinktą tam tikrą artistinį stilių. Šio GAN pranašumai yra tai, kad juo naudojantis galima išgauti nuotraukų stilius ir juos pritaikyti nemokinant generatyvinio besivaržančio neuroninio tinklo kiekvieną kartą iš naujo. Taip pat generuojant veidą galima nurodyti ir stiliaus intensyvumą nustačius indikatorius įvestį.

2.2.3 StyleGAN analizė

StyleGAN, kuris jau buvo aptartas 2.1 skyriuje, kaip veidus generuojantis GAN. Šis GAN gali ne tik generuoti veidu, bet ir turi galimybę juos sujungti. Pačio StyleGAN veikimas nesikeičia atsižvelgiant į tai ar generuojamos nuotraukos, ar jos sujungiamos. Jei GAN naudojamas veidams sujungti, keičiasi įvesčių skaičius mokymosi metu. Taip pat generavimo metu yra naudojamos dvi įvesties nuotraukos iš kurių yra gaunami veido bruožai, kurie vėliau yra sujungiami. Generavimo metu yra sujungiami pirmosios ir antrosios nuotraukos bruožai, taip

pakeičiant antros nuotraukos išvaizdą. Jei žiūrint į architektūrą (žr. 6 pav.) tai per atvaizdavimo tinklą (*angl. Mapping network*) įvedami du įvesties kodai z_1 , z_2 ir atitinkamai yra gaunami w_1 , w_2 vektoriai su kuriais ir yra valdomi veido bruožai. Taigi w_1 vektorius yra taikomas prieš vektorių w_2 , kur rezultate gaunamas bendras veidas iš dviejų nuotraukų [KLA19].

StyleGAN sugeneruoti rezultatai matomi 14 pav.



14 pav. StyleGAN veidų sujungimo pavyzdžiai [KLA19]

Nuotraukoje pavaizduoti StyleGAN sugeneruoti rezultatai naudojant šaltinius A ir B. Iš B šaltinio buvo imami veido stiliai: akiniai, plaukai, amžius, poza, lytis. A šaltinyje yra pateikti veidai, kurie bus keisti pagal šaltinį B. Iš rezultatų galima pastebėti, kad sugeneruoti rezultatai pakeičia šaltinio A nuotrauką į visiškai kitokio žmogaus veidą.

StyleGAN gali sugeneruoti ne tik veidus atsitiktinai pasirinkus tam tikrus stilius, bet taip pat gali realistiškai sujungti veidus tarpusavyje.

2.2.4 Apibendrinimas

Šiame poskyryje buvo išanalizuoti GAN, kurie pateikia skirtingus būdus, kaip galima sugeneruoti veidus atsižvelgiant į kito žmogaus veido bruožus. StyleGAN ir InterFaceGAN generuoja veidus atsižvelgiant į panašius veido atributus tokius, kaip poza, aksesuarai, lytis,

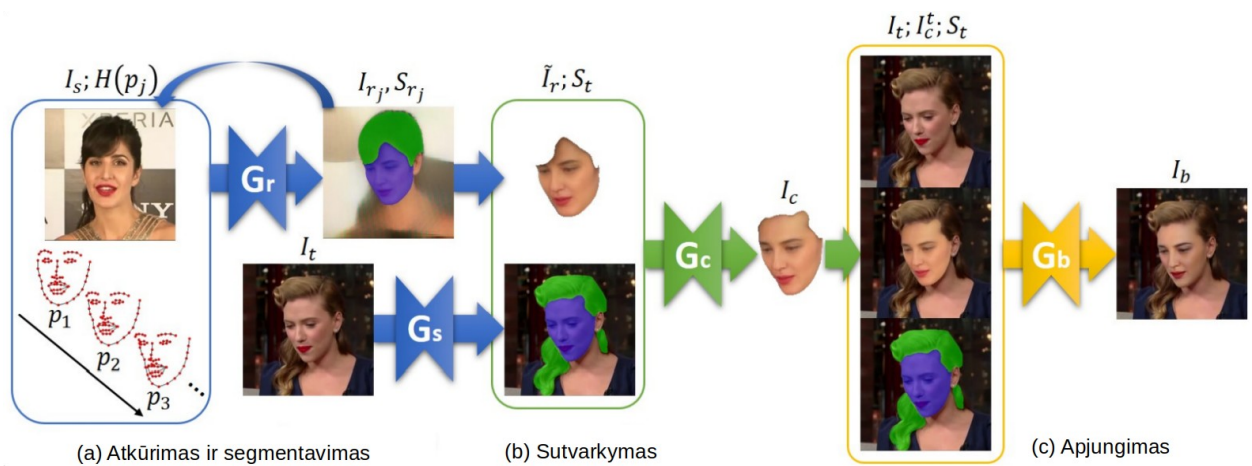
amžius. Bet abiejų GAN pateikiami sprendimai yra skirtingi. StyleGAN naudoja savo modifikuotą mokymo ir generavimo modelį. Tas pats modelis yra naudojamas ir veidų generavimui bei sujungimui, tik norint gauti skirtingą rezultatą reikia pakeisti įvesties kiekį. Norint sujungti veidus reikia pateikti dvi nuotraukas. Iš jų yra išgaunami veidų bruožai, kurie yra maišomi tarpusavyje, kol išgaunamas naujas veidas. InterFaceGAN pateikia būdą, kaip galima iš latentinės erdvės atsirinkus atsitiktines nuotraukas, sugrupuoti į kategorijas pagal bruožus ir pagal tuos bruožus sudėti ribas erdvėje. Latentinėje erdvėje keičiant pateiktos nuotraukos poziciją, keičiasi veido bruožai. BlendGAN pritaiko dar kitokį požiūrį sujungiant veidus. Šis modelis pasiūlė naują būdą, kaip galima pateiktą veidą stilizuoti pritaikant norimą stilizuotą meninį stilių, naudojant kitą stilizuotą veidą. Kur iš stilizuoto portreto yra paimamas, tik meninis stilius ir pritaikomas norimam veidui, pagal nurodytą indikatorius (stiliaus intensyvumą).

2.3. GAN skirtų veidų keitimui vaizdo įrašuose analizė

Šiame poskyryje bus apžvelgiami GAN, kurie keičia žmonių veidus vaizdo įrašė atsizvelgiant į tai, kokio žmogaus veidą norima pakeisti. GAN, kurie tai daro yra FSGAN, DeepFaceLab. Šie generatyviniai modeliai pateikia skirtingus sprendimus pakeičiant vieno žmogaus veidą kitu vaizdo įrašė.

2.3.1 FSGAN analizė

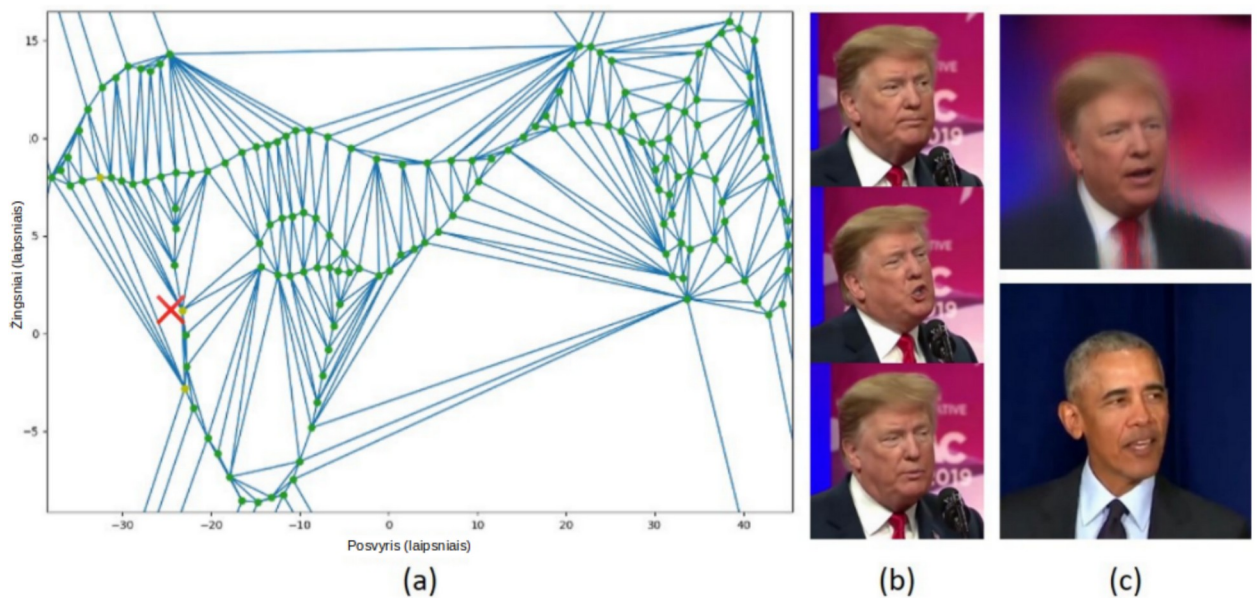
FSGAN pateikia sprendimą, kaip galima pakeisti vieno žmogaus veidą kitu veidu vaizdo įrašė neapmokant GAN modelio kiekvieną kartą. Pavyzdžiui pateikiant savo veidą, kad vaizdo įrašė pakeistų kokio nors esančio aktorius veidą ir po to pateikiant kito žmogaus veidą, kad pakeistu to pačio aktorius veidą, nereikia mokinti atskirai GAN modelio, tam atlikti. Taigi kaip yra atliekamas dinamiškas veido pakeitimas vaizdo įrašė be apmokymų realiu laiku: pirmiausiai paimamas vaizdo įrašas, kuriame bus keičiami veidai, toliau vaizdo įrašas yra suskirstomas į dalis su veido orientyrais ir mažais pasikeitimais veide. Po to yra paimamas veidas, esantis vaizdo įrašė ir pakeičiama jo orientacija į norimo veido atvaizdo orientaciją, pagal tai yra surandami abiejų veidų bruožai ir susegmentuojami. Pagrindiniai elementai, kurie būna susegmentuojami tai yra veido bruožai ir plaukai. Turint susegmentuotus vaizdus jie yra sujungiami ir neatitikimai tarp veidų yra užpildomi. Visus šiuos veiksmus galima pamatyti 15 pav. [NKH19].



15 pav. FSGAN veikimo pavyzdys [NKH19]

Paveiksle pavaizduotas bendro FSGAN veikimo principas, kaip sujungiami veidai be jokių apmokymų. Kaip ir minėta (a) punkte yra pavaizduotas I_s vaizdo įrašas iš kurio yra išgaunami veido pasikeitimai į p_1, p_2, \dots, p_n dalis. Toliau iš to pačio vaizdo įrašo yra išgaunamas veidas, kuriame ir pakeičiama pozicija taip kaip atrodo I_t nuotraukoje naudojant G_r generatorių. Po to yra susegmentuojami abiejų veidų – vaizdo įrašo ir norimo veido, veido bruožai ir plaukų dalys naudojant G_s generatorių. Toliau (b) punkte turint paruoštus segmentus ir atskirtą veidą, šie elementai yra perduodami į generatorių G_c . G_c generatorius sutvarko veido dalis, ten, kur trūksta veido dalies ją sugeneruoja, o kur yra per daug – nukerpa. Paskutiniame punkte (c) gautas sutvarkytas veidas yra paduodamas į generatorių G_b , kuris sujungia ir pritaiko veidą, kad atitiktų pagal suteiktą veido atvaizdą – atspalviai, apšvietimas ir t.t. Taigi pasinaudojant visais šiais žingsniais FSGAN gali realiu laiku pakeisti vieno žmogaus veidą kitu ir gauti vaizdo įrašą su kito žmogaus veidu.

Taip pat FSGAN pateikia išvaizdos žemėlapi (angl. *Appearance map*), kuriame yra saugoma vaizdo įrašo veido pozos ir veido bruožų pasikeitimai. Visa tai galima pamatyti 16 pav.

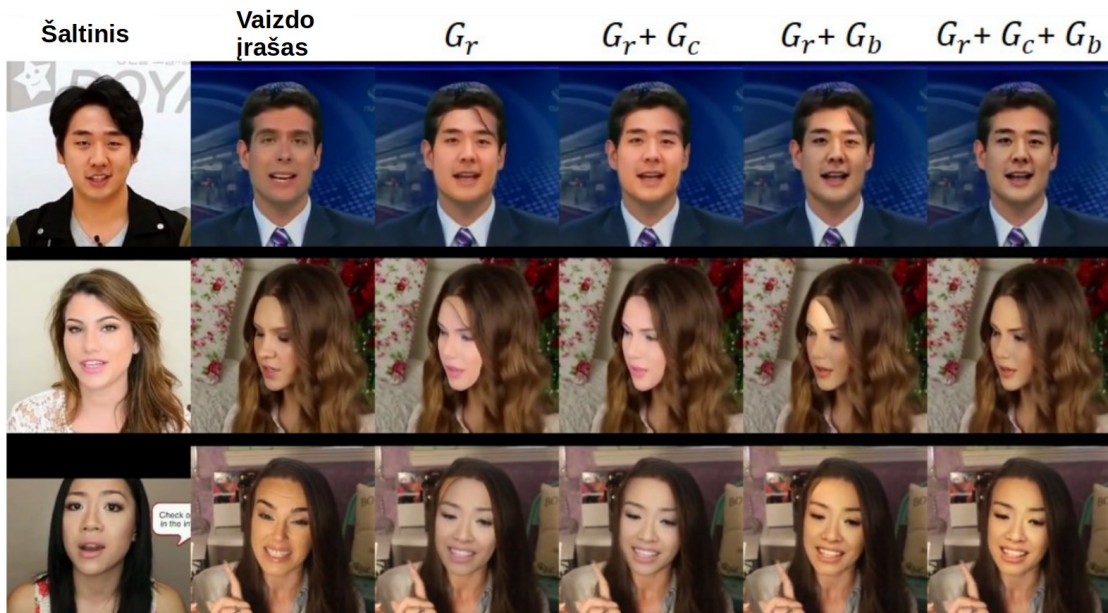


16 pav. FSGAN išvaizdos žemėlapis [NKH19]

Nuotraukoje matomas išvaizdos žemėlapis, kuriame pavaizduoti trys punktai:

- (a) Pункte pavaizduotas išvaizdos žemėlapis, kuriame yra atvaizduota šaltinio (b) subjekto pavyzdys. Žali taškai žymi šaltinio veido skirtingus atvaizdavimus, mėlynos linijos vaizduoja „Delaunay Triangulation“, tarp žalių taškų sujungimą, o raudonas X žymi dabartinę (c) punkto veido pozą;
- (b) Kaip ir minėta rodo šaltinio veidą, pagal kurį atliekami pakeitimai;
- (c) Rodo norimo užkeisti veido pasikeitimai ir originali veido nuotrauka.

Iš šio žemėlapio galima matyti pasikeitimus veido pozose, perėjimus tarp jų bei pastebėti, kaip gaunamas bendro veido pasikeitimo veiksmas. Toliau yra pateikiamas pavyzdys, kaip atrodo sugeneruoti veidai vaizdo įrašuose pereinant per kiekvieną FSGAN generatorių kol išgaunamas rezultatas 17 pav. [NKH19].



17 pav. FSGAN sugeneruoti pavyzdžiai [NKH19]

Paveiksle pavaizduota, kaip keičiasi vaizdo įrašo veidas, naudojant šaltinio veidą po kiekvieno generatoriaus rezultato. Nuo veido kuriam trūksta dalies ar neatitinka veido spalvos iki pačio galutinio rezultato kai viskas yra sutvarkyta ir gautas galutinis rezultatas.

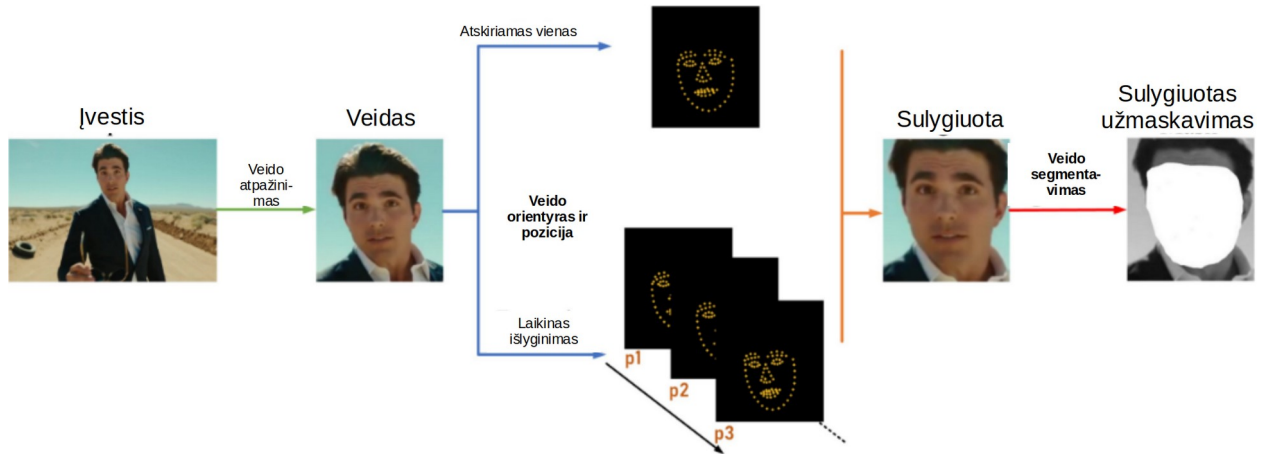
Išanalizavus FSGAN veikimą, galima matyti, kad pateikti sprendimai, leidžiantys vaizdo medžiagoje pakeisti vieną veidą kitu suteikia unikalų pranašumą. Ypatingai, kai tai galima atlikti realiu laiku ir kiekvieną kartą nemokant GAN modelio atskirai su dviejų žmonių veidais, taip išvengiant mokymo dalies.

2.3.2 DeepFaceLab analizė

DeepFaceLab yra atviro kodo projektas, kuris siekia padaryti lengvai pasiekiamą ir mažai resursų prašantį modelį, kuris sugeneruoja realistiškus veidų pakeitimus vaizdo įrašuose. Kuo skiriasi DeepFaceLab nuo FSGAN – kad išgautų geros kokybės veidų pakeitimą DeepFaceLab turi kiekvieną kartą apmokyti savo modelį iš naujo su norimu vaizdo įrašu ir veido atvaizdu, kuri reikės pritaikyti vaizdo įrašė. Norint sugeneruoti vaizdo įrašą naudojant DeepFaceLab yra atliekami trys pagrindiniai žingsniai:

1. Veido išgavimas;
2. GAN modelio mokymas vyksta dviem etapais, pirmiausiai pritaikant DF metodą, kuris išgauna veido detales, o po to pritaikomas *LIAE* metodas, kuris išgauna vaizdo įrašo apšvietimo informaciją;
3. Veidų sujungimas;

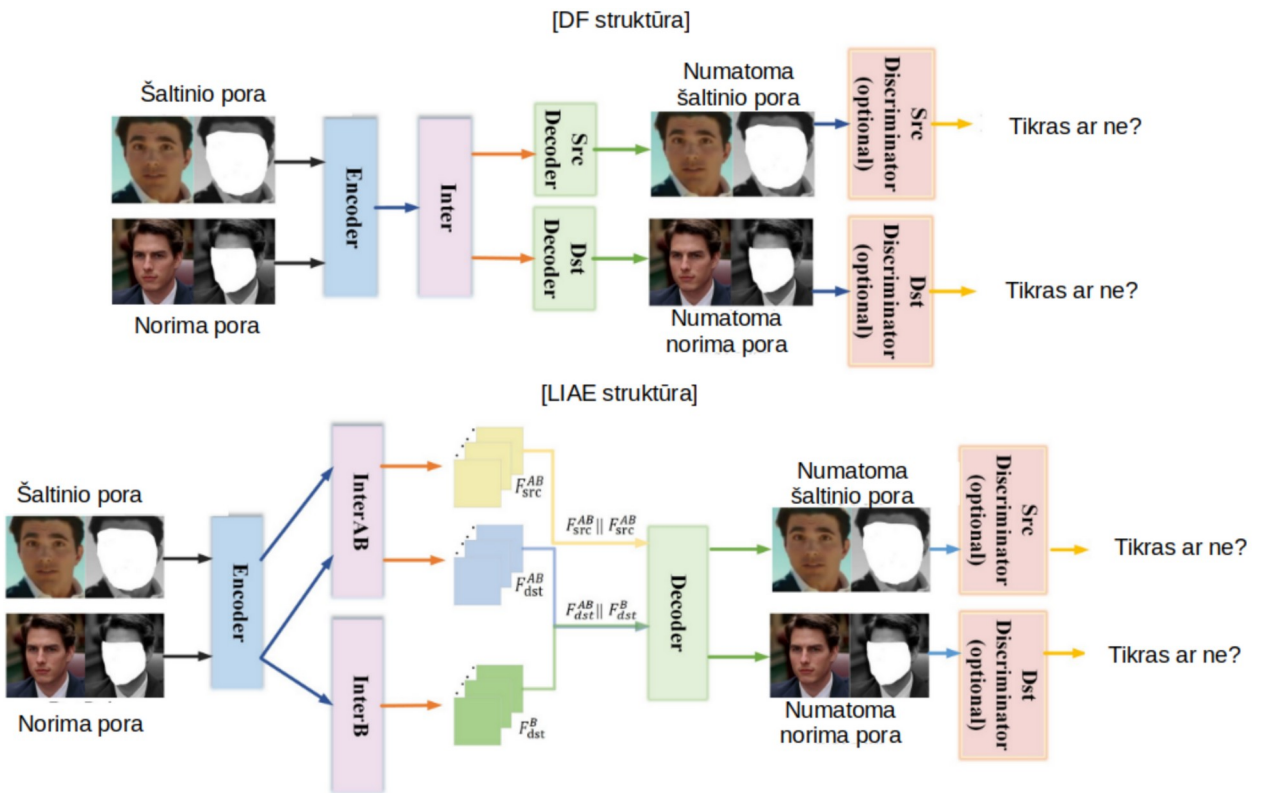
Veido išgavimas yra šiek tiek panašus į FSGAN veido išgavimą, kurio metu yra atrandamas veidas ir iš jo yra ištraukiami judesiai ir veido bruožai. Tai galima pamatyti 18 pav. [PGC+21].



18 pav. DeepFaceLab veido išgavimo pavyzdys [PGC+21]

Paveiksle matomas veido išgavimo procesas. Jis prasideda nuo pateikto vaizdo įrašo, toliau atliekamas veido atpažinimas ir po veido atpažinimo atskiriamas vienas viso veido kadras. Turint vieną viso veido kadrą pagal jį yra atpažįstamas tas pats veidas kituose kadruose ir tuomet veido pokyčiai suskaidomi į dalis p_1, p_2, \dots, p_n . Suskaidžius veidą į dalis, kiekviename kadre veidas yra sulygiuojamas, kad būtų matomas iš priekio arba šono ir galiausiai turimas veidas yra susegmentuojamas. Šis procesas sudaro duomenų rinkinį, iš kurio vėliau generatyvinis besivaržančių neuroninių tinklų modelis mokysis. Duomenų rinkiniai yra sudaromi su norimu vaizdo įrašo veidu bei veidu, kuris jį pakeis.

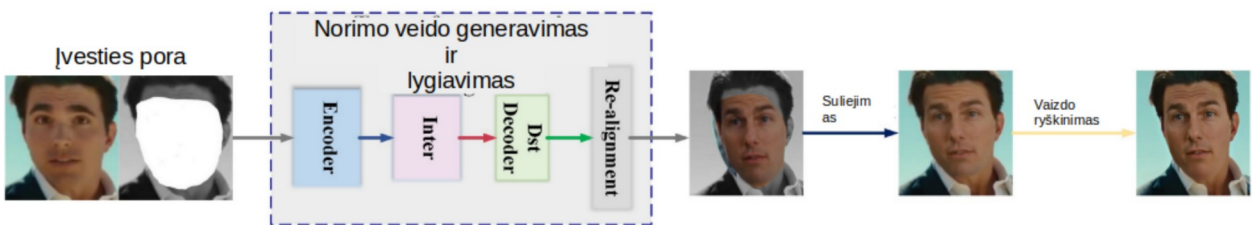
Atlikus visus šiuos veiksmus ir turint duomenų rinkinius yra atliekamas mokymas su *DF* ir *LIAE* metodais. *DF* metodas išgauna iš abiejų veidų bendrus bruožus ir juos sulygina, tačiau šis metodas nesugeba išgauti pozicijos, kaip turi pasidengti veidas ar tokių elementų kaip apšvietimo ir atspalvių vaizdo įrašė. Todėl tam išspręsti yra naudojamas *LIAE* metodas, kuris išgauna apšvietimo ir atspalvių detales. Šių dalių schemas galima pamatyti 19 pav. [PGC+21].



19 pav. DeepFaceLab mokymosi pavyzdys [PGC+21]

Paveiksle matoma *DF* ir *LIAE* metodų struktūros. *DF* metodas, turintis mažiau detalių struktūroje, su *Encoder* ir *Inter* funkcijomis, kurios atlieka pagrindinius veiksmus. *LIAE* turi panašią struktūrą, tik yra atskirai papildoma *Inter* funkcijos, kur į vieną paduodami abiejų veidų svoriai, o paskui kryžmiškai gražinamas atgal ir sujungiamas prieš perduodant į dekoduojoją. Taip išgaunant daugiau informacijos apie veidus ir kaip juos reikia sujungti.

Turint apmokytą modelį galima sujungti veidus ir gauti rezultatą. Veidų sujungimo schemą galima pamatyti 20 pav.



20 pav. DeepFaceLab veido generavimo pavyzdys [PGC+21]

Paveiksle pavaizduota, kaip sugeneruojamas naujas veidas. Iš pradžių yra pateikiama įvesties pora – norimas veidas ir jo segmentuota veido dalis. Toliau įvestis perduodama į tikslinio veido generatorių ir lygiavimą, kur nuotrauka yra užkoduojama, sutvarkoma ir pritaikoma vaizdo

įrašė. Po pritaikymo, veidas yra suliejamas, t. y. sutvarkoma forma, atspalvis ir po to gautas rezultatas yra paryšklinamas.

Toliau pateikiami pavyzdžiai, kaip atrodo galutinis rezultatas naudojant DeepFaceLab 21 pav.



21 pav. DeepFaceLab veido generavimo pavyzdys [PGC+21]

Paveiksle matomi rezultatai sugeneruoti DeepFaceLab. Pirmoji nuotrauka blokuose yra įvesties veidas, sekanti yra veidas, kurį reikia pritaikyti vaizdo įrašė. O paskutinė nuotrauka nurodo, kaip atrodo galutinis sugeneruotas rezultatas. Iš nuotraukos galima matyti, kad sugeneruoti rezultatai yra gana aukštos kokybės su mažai vyraujančių artefaktų ar klaidų.

Taigi DeepFaceLab turi kitokį būdą nei FSGAN, kaip pakeisti veidus vaizdo įrašė. Pagrindinis skirtumas šio generatyvinio besivaržančio neuroninio tinklo yra tai, kad norint išgauti vaizdo įrašė su pakeistu veidu pirmiausia reikia apmokyti patį GAN modelį ir tik po to generuoti galutinį rezultatą. Nors generavimo procesas yra ilgesnis, bet gaunami rezultatai yra aukštos kokybės, kaip tai galima matyti iš pavyzdžių.

2.3.3 Apibendrinimas

Išanalizavus pasirinktus GAN, galima pastebėti, kad nurodyti GAN sprendžia visiškai kitokias problemas generuojant veidus vaizdo įrašuose. FSGAN siūlo būdą, kaip galima pakeisti veidus vaizdo įrašė neapmokant GAN modelio kiekvienam atvejui atskirai. Šis modelis geba realiu laiku apdoroti duotą nuotrauką ir vaizdo įrašė. DeepFaceLab kitaip atlieka veido pakeitimą vaizdo įrašuose – iš pradžių yra pateikiamas duomenų rinkinys, kuris yra sudarytas iš norimo veido nuotraukų. Toliau pateikiamas vaizdo įrašas DeepFaceLab modeliui, kuris sukuria dar vieną veidų duomenų rinkinį iš pateikto vaizdo įrašo. Tuomet turint du duomenų rinkinius yra apmokomas GAN modelis ir tik tada yra generuojamas galutinis rezultatas.

2.4. Analizės apibendrinimas

Išanalizavus pasirinktus generatyvinius besivaržančius neuroninius tinklus parodė, kad GAN turi įvairių galimybių veido generavime ir atitinkamai – kad kiekvienas GAN turi savo stiprybes ir silpnybes, priklausomai nuo naudojamo modelio. StyleGAN2 turi galimybę valdyti veido stilių ir generuoti aukštos kokybės rezultatus, BigGAN generuoja realistiškus ir įvairius vaizdus, tačiau generuojant veidus turi didelę galimybę gauti veidus su iškraipymais ir artefaktais. PG-GAN suteikia galimybę generuoti aukštos kokybės nuotraukas, pakeisdamas GAN mokymo principą. Pereinant prie veidų sujungimo, InterFaceGAN pateikia unikalų būdą, kaip galima sujungti veidus parinkus norimus bruožus nekeičiant kitų, kai StyleGAN sujungia visus, o jei reikia pakeisti tik kelis veido bruožus reikia daugiau eksperimentavimo. BlendGAN pateikia kitokią galimybę keisti veidus – sujungiant piešto veido stilių su norimu veidu. Vaizdo įrašė veido keitimas pateikiamas skirtingais būdais: apmokant GAN modelį prieš generuojant vaizdo įrašą ir nereikalaujant GAN modelio mokymo.

3. Generatyvinių besivaržančių neuroninių tinklų eksperimentinė patikra ir vertinimas

Šiame skyriuje bus vertinami GAN modelių generuojami rezultatai pagal kategorijas, kur iš kiekvienos kategorijos bus parenkamas geriausias GAN modelis. Pats vertinimas bus atliekamas atsižvelgiant GAN unikalias galimybes ir į kiekybinius vertinimus naudojant: „Inception Score“ toliau bus vadinamas IS, „Frechet Inception Distance“ toliau bus vadinamas FID ir „Structural Similarity Index Measure“ toliau bus vadinamas SSIM. Šie vertinimai bus pateikiami gautais rezultatais, kuriuose iš kiekvieno GAN bus sugeneruoti po 1000 nuotraukų ir toliau paskaičiuojami IS ir FID rezultatai, taip pat bus sulyginami su straipsniais, kuriuose patys autoriai pateikia vertinimus. Taip pat šiame skyriuje bus pateikiami vertinimo metodai ir ką jie vertina bei GAN modelių įvertinimai ir palyginimai.

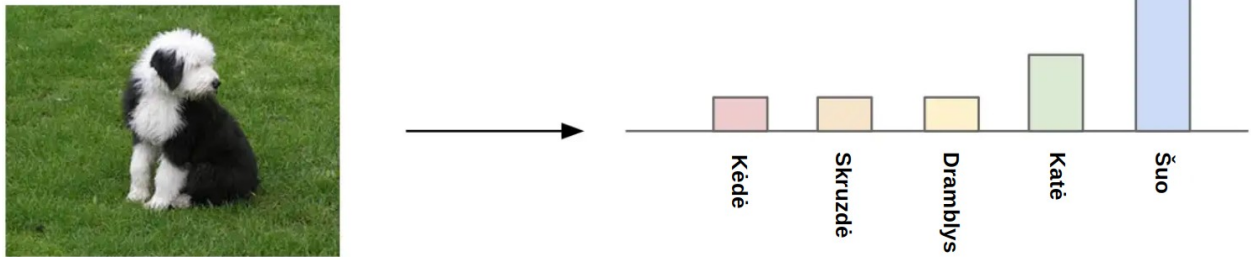
3.1. GAN vertinimo metodai IS, FID ir SSIM

Šioje dalyje bus apžvelgti minėti vertinimo metodai, t.y. IS, FID ir SSIM, kuriais naudojantis bus lyginami GAN modelių rezultatai. Šie metodai buvo pasirinkti, nes straipsniuose, tai yra vieni iš pagrindinių nesubjektyvių palyginimo metodikų, kurias naudoja patys autoriai. Bei šie metodai leidžia pamatyti kiekybinius rezultatus, kurie parodo GAN modelio galimybes.

Bus pradama nuo IS vertinimo metodo, kuris matuoja pagrindinius du kriterijus:

1. **Nuotraukų kokybė**, tai yra ar nuotrauka atspindi prieš tai buvusios nuotraukos tipą, pvz.: vienoje nuotraukoje buvo katė ir kituose nuotraukoje yra randamos katės;
2. **Nuotraukų įvairovė**, tai yra kai generuojamos nuotraukos yra įvairios nepaisant to, kad yra kuriamos to pačio tipo nuotraukas pvz.: generuojant kačių nuotraukas ar kiekviena sugeneruota nuotrauka turi vis kitos veislės katę.

Jei GAN modelio generuojamos nuotraukos tenkina šiuos abu kriterijus reiškiasi IS rezultatas bus aukštas, teoriškai IS maksimumo reikšmės nėra. Jei generuojamos nuotraukos netenkina kriterijų IS rezultatas bus žemas, mažiausia IS reikšmė gali būti 0. Taigi kuo didesnis IS rezultatas tuo daugiau įvairovės turi GAN modelis, kur gali sugeneruoti atpažįstamas ir bendros kategorijos nuotraukas, kaip katės, šunys, veidai ar kiti vaizdai [SGZ+16]. Visas rezultatas gaunamas pagal nuotraukų pasiskirstymą. Kaip nuotraukos pasiskirsto ir kaip jos yra atpažįstamos, toliau pateikiamas pavyzdys 22 pav.



22 pav. IS pasiskirstymo pavyzdys medium.com

Paveiksle matoma sugeneruota nuotrauka iš kurios nustatoma, kas ten matoma ir priskiriama kategorijai pvz.: šuo. Pagal dešinėje pusėje pateiktą grafą matoma, kad daugiausiai suklasifikuotų nuotraukų buvo šuns. Tai reiškia, kad generuojamos nuotraukos yra pagrinde atpažįstamos kaip šunys ir tai reikštų aukštesnį IS rezultatą, nes tai nusako, kad generuojamos nuotraukos yra kokybiškos ir atpažįstamos, kaip tos pačios klasės. Nes jei grafas būtų išsiskirstęs tolygiai, tai reikštų mažesnį IS rezultatą, nes nuotraukos nėra geros kokybės, taip nesuprantant bendrosios generuojamos kategorijos. IS problema yra tai, kad vertina bendrą nuotraukų kokybę ir įvairovę, tačiau nevertina ar tos nuotraukos yra realistiškos.

Realistiškumo balą suteikia FID vertinimo metodas. Jis vertina sugeneruotų vaizdų kolekciją su realių vaizdų kolekcija iš generuojamos srities. Šis vertinimas, kartu su IS vertinimu leidžia nusakyti ar GAN modelis generuoja geros kokybės, įvairius ir tikroviškus vaizdus. FID yra apskaičiuojamas lyginant sugeneruotos nuotraukos ir realaus vaizdo vektorių atstumus pagal generuojamą kategoriją ir taip įvertinant tikroviškumą. Kuo mažesnis įvertinimas tuo sugeneruota nuotrauka yra tikroviškesnė, o kuo FID rezultatas yra didesnis tuo generuojama nuotrauka yra mažiau realistiška [NKS21][HRU+17].

Paskutinis vertinimo metodas yra SSIM, kuris bus naudojamas vaizdo įrašų vertinime. Šis metodas vertina panašumus tarp vaizdo įrašo kadrų. Kur gautas rezultatas gali padėti nustatyti, kaip gerai pavyko sugeneruoti vaizdo įrašą. Bendras SSIM vertinimas susidaro iš šių parametrus:

- **Kontrasto;**
- **Šviesumo;**
- **Struktūros.**

Pagal šiuos tris parametrus yra įvertinami kadrai tarpusavyje ir pagal tai gaunamas atsakymas gali būti tarp -1 ir 1, kur -1 nusako, kad vaizdai tarpusavyje yra skirtingi, o 1 nusako,

kad vaizdai tarpusavyje yra labai panašūs. Įvertinimas $[0, 1]$, taip pat tinka ir nusako tą patį, kaip ir $[-1, 1]$ [NM20].

3.2. GAN įvertinimų palyginimai

Taigi išsiaiškinus vertinimo metodus, kaip ir ką jie vertina bus pateikiamos vertinimo lentelės pagal kategorijas. Rezultatai bus pateikiami iš straipsnių, kuriuos pateikia patys autoriai bei iš sugeneruotų rezultatų, kuriems bus naudojami straipsniuose pateikti apmokyti modeliai, išskyrus DeepFaceLab modelį, kurį reikės apmokyti. GAN modelių rezultatams generuoti yra naudojamas GTX 1060 6GB vaizdo plokštė.

3.2.1 GAN įvertinimas pagal veidų generavimą

Šioje dalyje bus pateikiami įvertinimai StyleGAN2, BigGAN ir PG-GAN veidų generavime pagal IS ir FID vertinimo metodus. Vertinimai gali būti matomi 1 lentelėje.

1 lentelė. GAN palyginimas, kurie generuoja veidus pagal IS ir FID

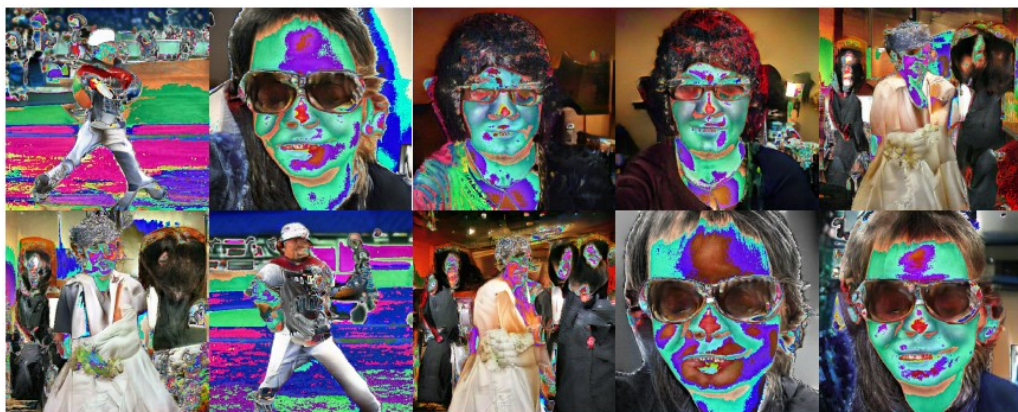
GAN	Autorių gauti rezultatai		Sugeneruotų veidų gauti rezultatai	
	IS	FID	IS	FID
StyleGAN2	5,17	2,70	7,54	31,69
BigGAN	166,5	7,4	5,97	212,77
PG-GAN	8,8	7,3	7,38	186,79

1 lentelėje matomi GAN modelių vertinimai pagal IS ir FID. Vieni rezultatai yra paimti iš straipsnių, kuriuose šie modeliai buvo sukurti [BDS19][KLA+20][KAL+18][KSP22], antrieji rezultatai yra gauti iš sugeneruotų nuotraukų, kuriems sugeneruoti buvo naudojami tie patys generatyviniai modeliai. Apžvelgiant rezultatus galima pastebėti, kad straipsniuose pateikti vertinimai skiriasi nuo gautų. Straipsnių gautuose rezultatuose geriausias IS vertinimas buvo apskaičiuotas naudojant BigGAN, kadangi IS nusako generuotų nuotraukų įvairovę, o BigGAN modelis pasižymi generuojamų nuotraukų kategorijų gausa. Lyginant pagal FID realistiškiausias nuotraukas generavo StyleGAN2. O pagal gautus rezultatus galima matyti, kad didžiausias IS bei FID vertinimas yra gaunamas su StyleGAN2, kas nusako jog šis generatyvinis besivaržantis neuroninis tinklas šiuo atveju generavo daugiausiai įvairovės turinčias bei realistiškiausias nuotraukas lyginant su kitais GAN. Taip pat galima pastebėti, jog gauti rezultatai skiriasi nuo pateiktų straipsnyje. StyleGAN2 atveju IS vertinimas yra geresnis, nes galimai straipsnyje [KLA+20] skaičiuojant rezultatus buvo imamas didesnis kiekis sugeneruotų nuotraukų, kuriose

galėjo pasitaikyti didesnis kiekis dublikatų. O FID yra prastesnis galimai dėl tos pačios priežasties, t. y., kad straipsnyje buvo naudotas didesnis nuotraukų kiekis. Analizuojant BigGAN gautus rezultatus, IS rezultatas yra ženkliai mažesnis lyginant su straipsnyje [BDS19] gautu, kadangi BigGAN modelis generuoja vaizdus, kurie gali būti priskirti tūkstančiams skirtingų kategorijų, o šiuo atveju visas sugeneruotas nuotraukas buvo galima priskirti nedideliamei kiekiui kategorijų, t. y. visoms kategorijoms, susijusioms su žmonėmis. FID rezultatas, kaip ir IS taip yra ženkliai prastesnis, kadangi BigGAN nėra sukurtas specifiskai žmonių veidų generavimui, todėl gauti rezultatai neatrodo realistiškai, kaip galima matyti 23 pav. Paskutinio modelio – PG-GAN vertinime galima išvelgti, jog IS yra gana panašus, tačiau FID skiriasi ženkliai. Tai galėjo atsitikti dėl to, kad didžioji dalis gautų nuotraukų turi nežymių artefaktų, kurie nors nėra labai akivaizdūs gali paveikti FID vertinimą.

Taigi atsižvelgiant į straipsnių rezultatus matosi, kad BigGAN sugeneruoja daugiausiai įvairovės palyginus su kitais GAN. Taip pat matoma, kad realistiškiausius rezultatus gauna StyleGAN2 toliau PG-GAN ir BigGAN tarpusavyje palyginus panašiai. Toliau atsižvelgiant į sugeneruotus rezultatus jie yra kitokie – StyleGAN2 generuoja įvairesnius vaizdus lyginant su kitais GAN ir gaunama, kad vis tiek yra realistiškiausi, po to seka PG-GAN ir BigGAN.

Toliau bus pateikiamos BigGAN modelio sugeneruotos nuotraukos 23 pav.



23 pav. BigGAN gauti rezultatai

BigGAN veidai buvo generuojami pagal kelias kategorijas kuriuose yra žmonės, kaip teniso žaidėjas, nuotaka, akiniai, perukas, nes specifinio veidų kategorijos neturi. Rezultatuose matosi kažkokie panašumai į generuotas kategorijas tačiau jie nėra realistiški, jei atsižvelgiama į veidų generavimą. Toliau pateikiami PG-GAN sugeneruoti rezultatai 24 pav.



24 pav. PG-GAN gauti rezultatai

PG-GAN rezultatai atrodo realistiškesni nei BigGAN generuojami, tačiau irgi turintys savo minusų, kaip neteisingai generuojami plaukai, artefaktai pačiose nuotraukose bei buvo sugeneruota nemažai dublikatų. Toliau pateikiami StyleGAN2 sugeneruoti rezultatai 25 pav.



25 pav. StyleGAN2 gauti rezultatai

StyleGAN2 rezultatai gauti yra realistiškiausi atsižvelgiant į rezultatus gautus su BigGAN bei PG-GAN. Taip pat StyleGAN2 rezultatai turi daugiau įvairovės ir mažiau artefaktų nuotraukose. Ne visos nuotraukos yra idealios, tačiau turi daugiau įvairesniu bei nereikia ilgai ieškoti tarp sugeneruotų nuotraukų, kol atrandami realistiški veidai.

Atsižvelgus į rezultatus galima matyti, kad veidams generuoti geriausias GAN būtų StyleGAN2, nes sugeneruoja realistiškiausias nuotraukas bei didžiausia įvairovę. Taip pat tai atsivaizduoja ir iš gautų IS bei FID rezultatų. Toliau lieka PG-GAN ir BigGAN, kur BigGAN veidų generavimui nėra tinkantis, nuotraukos neturi realistinių rezultatų. Tai lieka PG-GAN, kuris, generuoja veidus, kurie atrodo kaip žmonės tačiau yra daug panašių rezultatų bei rezultatai daugiau turi artefaktų nei StyleGAN2 generuojami rezultatai. Taip pat StyleGAN2 turi galimybę

valdyti veido stilių, tai yra bendrus veido bruožus, amžių, odos spalvą, suteikia didelį pranašumą norint sugeneruoti duomenų rinkinius ar specifinius veidus lengviau ir lanksčiau. Taip pat iš veidų generavimo tematikos buvo parašytas ir pateiktas straipsnis į *IEEE Access (Manuscript ID Access-2024-15725)* įkėlimo patvirtinimą galima matyti 1 priede.

3.2.2 GAN įvertinimas pagal veidų generavimą iš kelių nuotraukų

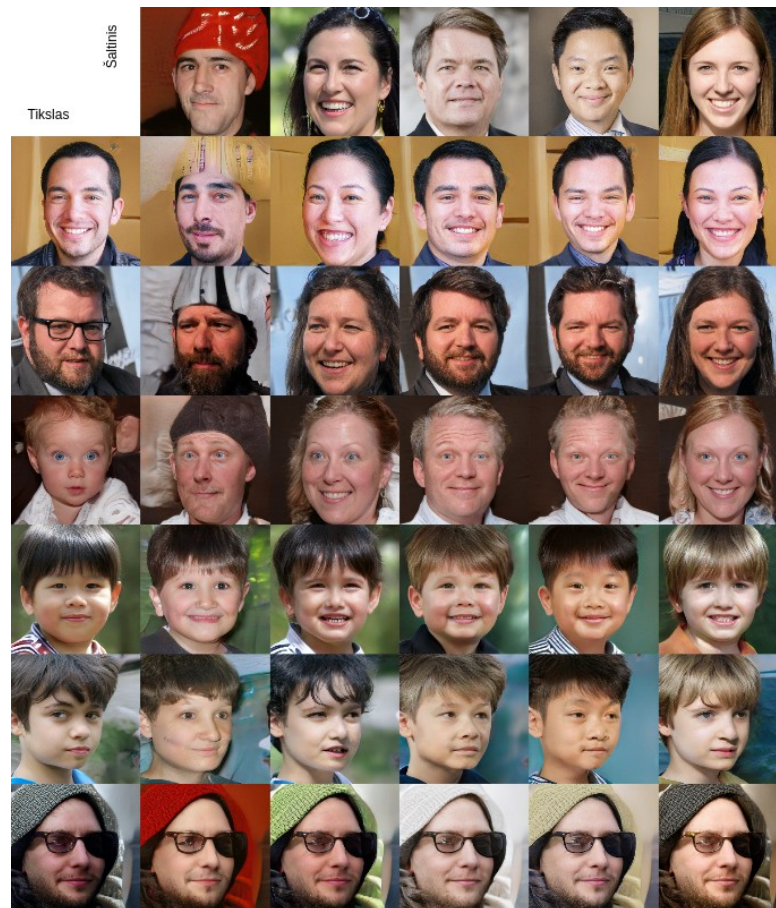
Šioje dalyje bus pateikiami BlendGAN, InterFaceGAN ir StyleGAN įvertinimai veidų generavime iš kelių nuotraukų pagal FID vertinimo metodą. IS vertinimo metodika nėra taikoma šiam generavimo tipui, kadangi generuojami veidai gali atrodyti vienodai keičiant smulkias detales ir dėl to IS vertinimas gali būti labai prastas, nes nuotraukose bus labai mažai įvairovės. Galimai dėl tos pačios priežasties nei vienas straipsnio autorius neturėjo IS vertinimo. FID rezultatas buvo apskaičiuojamas išvedus vidurkį iš gautų nuotraukų grupių. Grupės sudarė tokios nuotraukos, kuriose veidai buvo sugeneruojami atsižvelgiant į specifinius veido bruožus, tokius, kaip amžius, akiniai, lytis, poza, šypsena, t. y. Grupės atspindėjo veido bruožai. Šis būdas buvo taikomas StyleGAN ir InteFaceGAN, o BlendGAN naudoja kitokį veikimo principą, todėl veidai buvo generuojami su 30 skirtingų stilių, kuriuos pateikia kūrėjas ir iš jų, taip pat, kaip iš prieš tai minėtų modelių, buvo išvestas FID vidurkis. Toliau 2 lentelėje pateikiamas GAN veidų generavimo iš kelių nuotraukų palyginimas.

2 lentelė. GAN, kurie generuoja veidus iš kelių veidų, palyginimas pagal FID

GAN	Autorių gauti rezultatai	Sugeneruotų veidų gauti rezultatai
	FID	FID
BlendGAN (i = 0)	8,97	122,03
InterFaceGAN	-	77,58
StyleGAN	4,4	88,54

Lentelėje matomi įvertinimai pagal FID. Kaip ir buvusioje lentelėje iš vertinimų galima pastebėti, kad geriausi įvertinimai yra vėl gaunami iš straipsnių [KLA19][SGT+20][MQZ+21]. Pagal straipsnius StyleGAN turi geriausią rezultatą, toliau seka BlendGAN, o InterFaceGAN įvertinimo neturi. Atsižvelgiant į gautus rezultatus, InterFaceGAN rezultatai buvo realistiškiausi, galimai dėl to, jog šis modelis turi jau paruoštus veidų bruožų failus, kuriuos pritaikius pakeičia atitinkamai pateiktus veidus. Dėl to atsiranda galimybė sugeneruoti daugiau tikslesnių veido nuotraukų. Toliau seka StyleGAN ir BlendGAN. StyleGAN gavo prastesnius rezultatus nei InterfaceGAN galimai dėl to, jog su šiuo modeliu reikia daugiau eksperimentuoti, kol yra

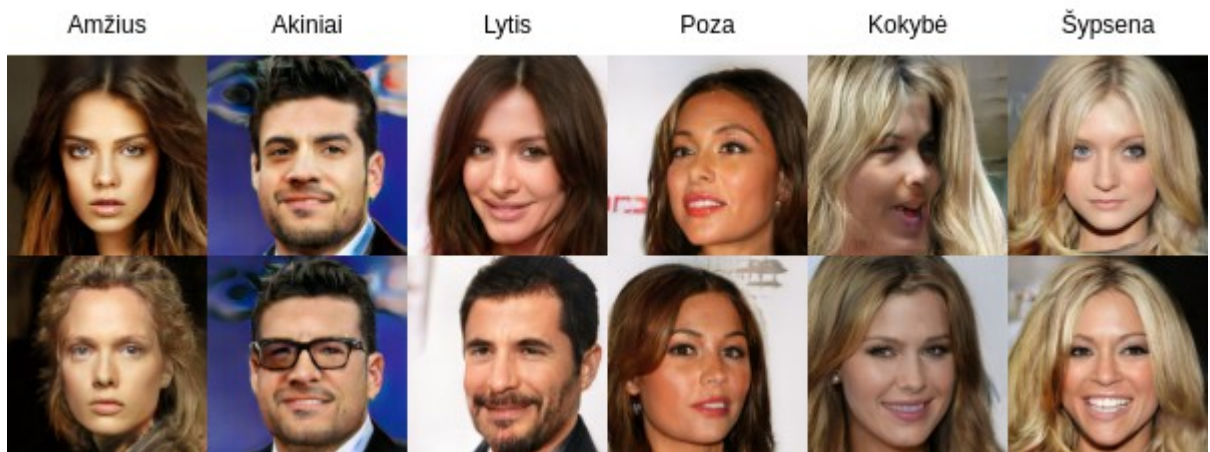
atrandami tam tikri veido bruožai, kuriuos reikia keisti, o atradus naudojimas panašus. Dėl to galimai galėjo pasitaikyti daugiau netikslių nuotraukų vertinimo metu. Paskutinėje vietoje atsidūrė BlendGAN, nes šis modelis generuoja stilistinius veidus, o FID vertina realistiškumą, tad šioje vietoje rezultatas buvo gautas, toks, kokį buvo galima numatyti. Galima teigti jeigu InterFaceGAN autorius būtų pateikęs FID skaičiavimus, jie būtų geresni arba labai panašūs su StyleGAN. Be to InterFaceGAN naudoja generuotus veidus, kuriuos toliau modifikuoja, o norint sugeneruoti nuotrauką dažniausiai yra naudojami StyleGAN arba PG-GAN modeliai, nors galima naudotis ir kitais veidus generuojančiais modeliais. Sugeneruoti nuotraukas, naudojantis BlendGAN galima pateikiant tiek sugeneruotus veidus, tiek realius, kuriems toliau yra nurodomas vienas ar keli stiliai ir galiausiai sugeneruojamas rezultatas su norimais pakeitimais. Toliau 26 pav. pateikiami StyleGAN rezultatai.



26 pav. StyleGAN gauti rezultatai

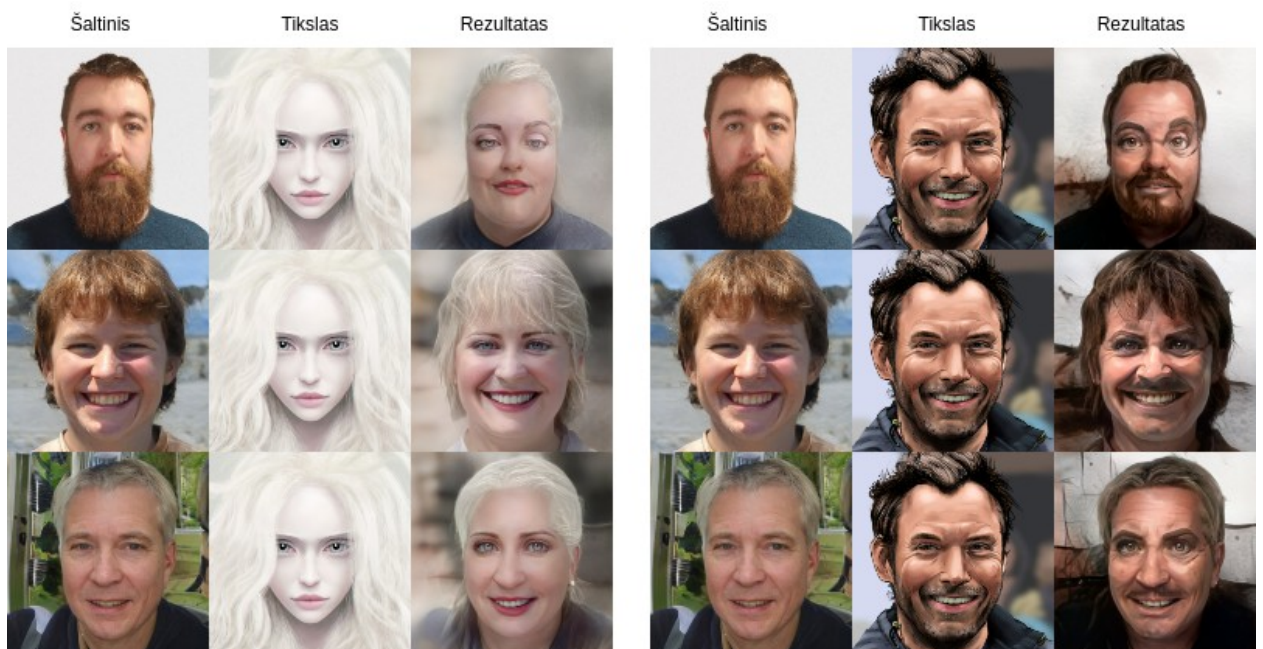
Paveiksle pateikiami StyleGAN sugeneruoti veidų jungimo rezultatai. Resultatuose tiek veidai šaltiniai (iš kokios nuotraukos yra generuojama), tiek veidai tikslai (į nuotrauką, kurią bandoma pakeisti) yra naudojami sugeneruoti. Pagal rezultatus galima matyti, kad ne visur

tikslas yra įgyvendintas, tačiau rezultatai yra vis tiek realistiški ir geros kokybės. Toliau pateikiami InterFaceGAN rezultatai 27 pav.



27 pav. InterFaceGAN gauti rezultatai

Paveiksle matomi InterFaceGAN rezultatai, kuriuose yra pakeistas amžius, uždėti akiniai, pakeista lytis, poza, pagerinta generuotos nuotraukos kokybė bei pridėta šypsena. Iš sugeneruotų vaizdų, galima matyti, kad rezultatus generuoja kokybiškus ir gana realistiškus. Toliau pateikiami BlendGAN sugeneruoti vaizdai 28 pav.



28 pav. BlendGAN gauti rezultatai

Pateiktuose BlendGAN rezultatuose galima matyti realaus veido nuotrauką, stilių, kuris buvo pritaikytas bei gautą rezultatą sujungus veidą su stiliumi. Pagal rezultatus, tos nuotraukos, kuriose asmenys yra vaizduojami su barzdomis BlendGAN nesugeneruoja tikslaus atvaizdo.

Barzda yra pratęsiama kaip veido dalis, o jei veidas neturi barzdos rezultatai yra geresni. Jei skiriasi realaus ir stiliuje vaizduojamo asmens lytis, rezultate bandoma paimti iš stiliaus veido bruožus bei juos pritaikyti realiam veidui, t. y., kad rezultate yra gaunamas ne tik pritaikytas stilius, bet ir pakeisti originalūs veido bruožai. Šie rezultatai buvo gauti pasirinkus identifikatoriaus reikšmę lygią 6, kas nusakytų, jog reikia pritaikyti tik stilių, tačiau būtų galima pabandyti pritaikyti dar mažesnę reikšmę, tuomet galimai būtų galima gauti rezultatus, kuriuose veido bruožai nebūtų keičiami.

Atsižvelgus į rezultatus bei generuojamas nuotraukas, galima matyti, kad InterFaceGAN generuoja realistiškiausius rezultatus bei naudojantis šiuo GAN iš karto yra pateikiami reikalingi failai specifiniams veido bruožams keisti. StyleGAN taip pat generuoja realistiškus rezultatus, tačiau pasitaiko ir tokių nuotraukų, kurios yra sugeneruojamos nekorektiškai arba neatsižvelgiant į numatyto veido bruožus. Be to norint gauti norimą rezultatą su StyleGAN reikia daugiau eksperimentuoti, nes šis generatyvinis modelis neturi jau pateiktų papildomų failų, kaip InterFaceGAN. Pagal FID vertinimą BlendGAN rezultatai yra patys nerealistiškiausi sulyginus su kitais apžvelgtais generatyviniais modeliais. Tačiau tai ir nėra šio GAN tikslas, nes BlendGAN sujungia realų pateiktą veidą su stilizuotu veidu. BlendGAN rezultatuose buvo matomi kokybiški rezultatai su mažai artefaktų turinčių vaizdų, tačiau ne su visais veido bruožais generatyvinis modelis susitvarko. Jeigu pateiktoje nuotraukoje žmogus turi barzdą, kaip galima matyti 28 pav. rezultatas yra iškraipomas ir gaunami veido bruožai galutiniam rezultate neatitinka bruožų pateiktų įvestyje.

3.2.3 GAN veidų generavimo vaizdo įrašuose palyginimas

Šioje dalyje bus pateikiami DeepFaceLab, FSGAN veidų generavimo vaizdo įrašuose įvertinimai pagal SSIM vertinimo metodą. Kaip ir buvo minėta vaizdo įrašams buvo pasirinktas SSIM vertinimas, nes FID ir IS vertinimo metodai yra taikomi nuotraukų vertinimui. SSIM vertinimo metodą rinkosi ir straipsnių autoriai [PGC+21] [NKH19]. 3 lentelėje pateikiami įvertinimai pagal SSIM metodą.

3 lentelė. GAN, kurie generuoja veidus vaizdo įrašuose, palyginimas pagal SSIM

GAN	Autorių gauti rezultatai	Sugeneruotų veidų gauti rezultatai
	SSIM	SSIM
DeepFaceLab	0,73	0,97
FSGAN	0,51	0,68

Pagal rezultatus pateiktus 3 lentelėje galima matyti, kad DeepFaceLab generuoja realistiškesnius vaizdo įrašus negu FSGAN, pagal autorių ir pagal gautus rezultatus. Tačiau, tai turėtų būti nenuostabu, nes norint sugeneruoti vaizdo įrašą naudojant DeepFaceLab pirmiausiai reikia sudaryti duomenų rinkinį norimam veidui pakeisti. Po to naudojant sudarytą duomenų rinkinį reikia apmokyti GAN modelį, kuris tik apmokius modelį sugeneruoja rezultatą. FSGAN gali sukeisti veidus nemokius GAN modelio kiekvieną kartą, todėl gali atsirasti didesnė netikslumų rizika, kadangi FSGAN nėra skirtas specifiniam veidui. Taigi geriausias iš apžvelgtų generatyvinių modelių veido keitimui vaizdo įrašė būtų DeepFaceLab dėl savo realistiškesnių rezultatų. Taip pat apskaičiavus sugeneruotų vaizdo įrašų rezultatus pagal SSIM ir palyginus su autorių gautais galima pastebėti, jog abiejų modelių rezultatai yra geresni. Tai galimai galėjo atsitikti dėl to, kad vaizdo įrašas kuris buvo vertinamas buvo 6 sekundžių ilgio ir tai galėjo daryti įtakos geresniam vertinimui, kadangi buvo mažiau kadrų. Toliau pateikiami DeepFaceLab ir FSGAN rezultatai 29 pav.



29 pav. DeepFaceLab ir FSGAN gauti rezultatai

Rezultate matoma, kaip sujungiamas šaltinio veidas su tikslo veidu. FSGAN išgauna gerą rezultatą, tačiau veido spalva skiriasi nuo originalaus, taip pat akys atrodo nenatūraliai palyginus su DeepFaceLab rezultatu. Taip pat DeepFaceLab gautas rezultatas geriau atvaizduoja veido mimikas nei FSGAN.

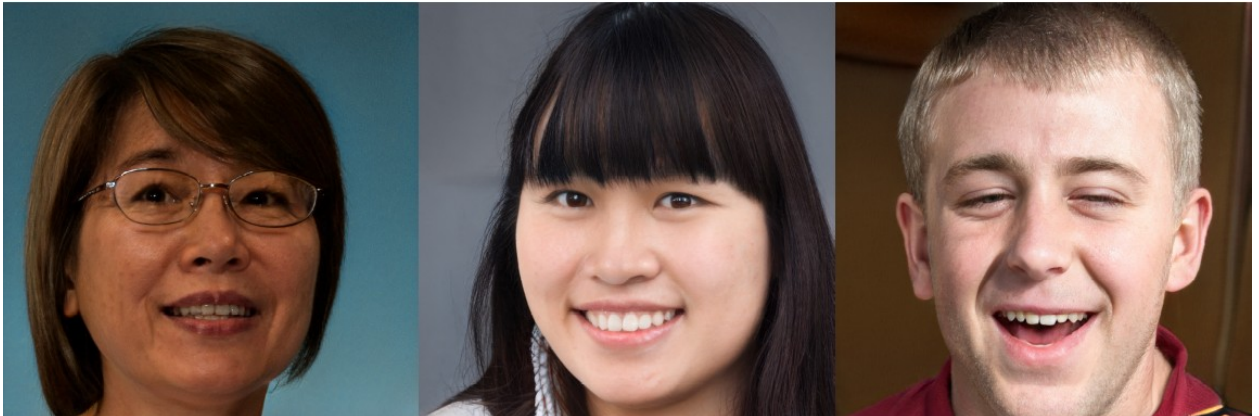
Apžvelgus FSGAN ir DeepFaceLab SSIM ir generuojamus rezultatus, galima teigti, kad DeepFaceLab generuoja geresnius rezultatus. Kaip ir buvo jau minėta DeepFaceLab gauna geresnius rezultatus dėl to, kad prieš sugeneruojant vaizdo įrašą su pakeistu veidu pirmiausia reikia apmokyti generatyvinį modelį, kuris tik vėliau gali pateikti rezultatą.

3.3. Eksperimentinės dalies apibendrinimas

Šioje dalyje buvo aptarti tokie generuotų nuotraukų vertinimo metodai, kaip IS, FID, kuriuos naudojant galima įvertinti generuojamų nuotraukų įvairovę, kokybę bei realistiškumą. O su SSIM vertinimu galima įvertinti generuojamo vaizdo įrašo kokybę.

Naudojant šiuos vertinimo metodus buvo palyginti StyleGAN2, BigGAN ir PG-GAN modelių sugeneruotos veidų nuotraukos. Pagal gautus rezultatus buvo galima pastebėti, kad geriausius vertinimus pavyko gauti StyleGAN2 generatyviniui modeliui, nors lyginant su straipsniais [BDS19][KLA+20][KAL+18][KSP22] gauti vertinimai išsiskyrė ir didžiausią įvairovę pavyko sugeneruoti BigGAN modeliui, o realistiškiausias nuotraukas, kaip ir gautuose rezultatuose, pavyko generuoti StyleGAN2.

Apžvelgiant aprašytus generatyvinius modelius vizualiai galima pastebėti, kad geriausius rezultatus taip pat generavo StyleGAN2, pavyzdys pateikiamas 25 pav. Nors šie rezultatai atrodo tikrai gerai, tačiau analizuojant sugeneruotas nuotraukas, galima rasti dar geriau atrodančių veidų, kurių nuotraukos yra pateikiamos 30 pav.



30 pav. StyleGAN atrinkti realistiški veidai

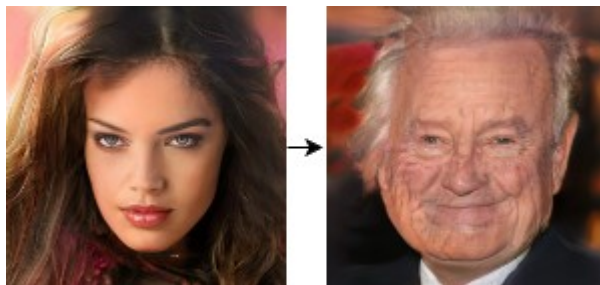
Kaip galima matyti 30 pav. pavaizduoti veidai yra sugeneruoti labai realistiškai, turintys mažai pastebimų artefaktų, kas sudaro įspūdį, kad šiose nuotraukose gali būti vaizduojami tikrų žmonių veidai. Tačiau, kad ir kaip realistiškai atrodančių nuotraukų galima rasti, pasitaiko ir tokių, kuriose akivaizdžiai matosi sugeneruoti artefaktai, jie iškraipo nuotraukos realistiškumą, kaip pavaizduota 31 pav.



31 pav. StyleGAN prastai sugeneruoti veidai

Remiantis 31 pav. galima pastebėti, jog šie vaizduojami žmonės yra sugeneruoti, kadangi yra sulietų detalių veiduose arba tokių bruožų, kurie natūraliai nepasitaiko. Be veido bruožų matosi ir plaukų struktūra bei fonas, kurie taip pat signalizuoja, kad tai yra generuota nuotrauka.

Naudojant FID vertinimą buvo įvertinti StyleGAN, InterFaceGAN bei BlendGAN modelių sugeneruotos veidų nuotraukos sujungiant kelis veidus. Geriausias rezultatus parodė InterFaceGAN lyginant gautus vertinimus iš sugeneruotų vaizdų, o lyginant straipsniuose gautus rezultatus – StyleGAN. Tačiau galima manyti, kad InterFaceGAN galėtų gauti geresnius rezultatus straipsnių vertinimo atžvilgiu, jei autorius būtų pateikęs įvertinimą. Taip pat galima matyti, jog InterFaceGAN generavo realistiškesnius vaizdus. Tačiau veidų sujungimo procese, kaip ir veidų generavimo procese ne visi rezultatai būna realistiški, todėl reikia pasirinkti tinkamus vaizdus. 32 pav. galima matyti netiksliai sujungtą veido pavyzdį, kuriame buvo bandyta pakeisti merginos amžių.



32 pav. InterFaceGAN prastai sugeneruoti veidai

Kaip ir buvo minėta, paveiksle matomas vienas iš blogai sugeneruotų pavyzdžių. InterFaceGAN bandė pasendinti veidą ir šio proceso metu pakeitė lyti bei visiškai pakeitė visus veido bruožus ir taip sugeneravo nerealistišką rezultatą.

Naudojant SSIM vertinimą buvo įvertinti DeepFaceLab ir FSGAN modelių sugeneruoti vaizdo įrašai sujungiant veidus. Pagal straipsnius ir gautus rezultatus realistiškiausius rezultatus generuoja DeepFaceLab. Tačiau norint išgauti geriausią rezultatą neužtenka tik apmokyti generatyvinio modelio, taip pat reikia nusistatyti tam tikrus parametrus tokius, kaip sujungimo tipas, šviesumas, spalvų kontrastas ir dar daugiau. Visi šie parametrai gali pakeisti galutinį rezultatą drastiškai 33 pav.



33 pav. DeepFaceLab prastai sugeneruotas kadras

Iš paveikslo galima matyti, kad sugeneruotame kadre pakeistas veidas yra ne tokios spalvos, kaip kakta, taip pat akių ir veido pozicija nėra pasukta į tinkamą pusę atsižvelgiant į galvos poziciją.

Galiausiai galima teigti, jog **StyleGAN2** yra tinkamas veidų generavimui, šio modelio gauti rezultatai yra realistiški, bet kaip ir kiti modeliai daro klaidų generuojant. Veidų sujungime geriausiai tai darė **InterFaceGAN**, nes šio modelio rezultatai buvo realistiškiausi bei generavimui yra paruošti reikalingi papildomi failai specifiniams veido bruožams pakeisti. Tačiau kaip ir su StyleGAN2 reikia patikrinti ir pasirinkti tinkamus rezultatus. Vaizdo įrašų generavime geriausiai pasirodė **DeepFaceLab**, jis generuoja realistiškiausius rezultatus, tačiau jį reikia apmokyti bei generuojant galutinį rezultatą reikia paeksperimentuoti su įvairiausiais parametrais norint išgauti realistiškiausią ir geriausią rezultatą.

Išvados

1. Apžvelgta GAN modelio struktūra, kuri susideda iš tokių dalių, kaip generatoriaus ir diskriminatoriaus. Generatorius gaudamas įvestį sugeneruoja naujus duomenis, o diskriminatorius tikrina, ar sugeneruoti duomenys yra realistiški. Taip pat buvo nustatyta, kad pritaikius GAN modelį galima atlikti tokius veiksmus, kaip veidų generavimas, veidų generavimas sujungus veidus ir veidų sujungimas vaizdo įrašuose, o šie specifiniai generavimo būdai gali būti taikomi tokiose srityse, kaip: kūryboje, akademinuose tyrimuose, žaidimuose, filmuose, IT srityje bei elektroninėje prekyboje. O, kad apžvelgti ir patikrinti galimus generavimo rezultatus buvo pasirinkti modeliai detalesnei analizei:
 - Veidų generavimui – StyleGAN2, PG-GAN bei BigGAN ;
 - Veido generavimui iš kelių kitų veidų ar nuotraukų – StyleGAN, InterFaceGAN, BlendGAN;
 - Veido generavimui ar keitimui vaizdo įrašė – DeepFaceLab ir FSGAN.
2. Atlikus išsamią analizę, buvo pastebėta, kad StyleGAN2 generuoja realistiškus veidus bei turi galimybę valdyti veido stilių, tačiau vis tiek pasitaiko nekokybiškų ar iškraipytų nuotraukų. BigGAN generuoja realistiškus ir įvairius vaizdus. Tačiau generuojant rezultatus turi didelę galimybę gauti sugeneruotų nuotraukų su artefaktais. PG-GAN suteikia galimybę generuoti aukštos kokybės nuotraukas, tačiau mokymo metu reikia pateikti aukštos kokybės nuotraukas bei išlaikyti jų poziciją ir vientisumą, kad nesugadinti mokymo proceso. Pereinant prie veidų sujungimo, InterFaceGAN pateikia unikalų būdą, kaip galima sujungti veidus pagal pasirinktą veido bruožą, nekeičiant kitų bruožų. Tačiau norint sujungti kelis veido bruožus, gauti rezultatai gali gautis nerealistiški arba turintys artefaktų. StyleGAN sujungia visus veido bruožus, tačiau jeigu reikia sujungti nuotraukas pasirinkus tik vieną veido bruožą, tam atlikti reikia daugiau eksperimentavimo. BlendGAN pateikia kitokią galimybę keisti veidus – sujungiant piešto veido stilių su norimu veidu nemokant GAN modelio kiekvieną kartą iš naujo. Tačiau jei norima sujungti tikrus veidus tarpusavyje BlendGAN to padaryti negali. Vaizdo įrašė veido keitimas pateikiamas skirtingais būdais: apmokant GAN modelį, bei neapmokant. Su DeepFaceLab prieš generuojant vaizdo įrašą reikia apmokyti generatyvinį modelį, taip gaunant realistiškesnį rezultatą bet šiam modeliui norint sugeneruoti realistišką rezultatą

reikia daugiau laiko ir resursų. O FSGAN nereikalauja apmokyti modelio kiekvieną kartą taip sutaupant laiko, tačiau rezultatai gali būti prastesni.

3. Atlikus eksperimentinius skaičiavimus ir palyginus rezultatus, galima teigti, kad veidų generavimui geriausiai tinka StyleGAN2. Atsižvelgiant į FID StyleGAN2 sugeneruoti vaizdai gavo apie 571 % geresnį vertinimą lyginant su BlendGAN, o palyginus su PG-GAN rezultatas buvo geresnis apie 489 %. Atsižvelgiant į IS vertinimo metodiką StyleGAN2 gauti rezultatai buvo geresni apie 26 % lyginant su BlendGAN ir apie 2 % lyginant su PG-GAN. Veidų generavimui iš kelių veidų geriausiai tinka InterFaceGAN. Pagal FID rezultatus gavo apie 57 % geresnį įvertinimą lyginant su BlendGAN ir apie 14 % lyginant su StyleGAN. Tačiau jei reikalingas stilistinis veido sujungimas, tuomet BlendGAN yra geresnis pasirinkimas, dėl kitokio generuojamo rezultato. Vaizdo įrašo veidų keitime geriausią įvertinimą pagal SSIM gavo DeepFaceLab apie 43 % geresnį rezultatą. Apskaičiuoti sugeneruotų nuotraukų rezultatai skyrėsi nuo autorių pateiktų, nes vertintas nuotraukų kiekis galėjo skirtis, kadangi ne visų modelių autoriai pateikė kiek nuotraukų naudojant buvo atliekami vertinimai. Taip pat rezultatas galėjo skirtis dėl naudotos ne vienodos įrangos ar skirtingų/naujesnių naudotų bibliotekų.

Literatūra

- [BWL20] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020.
- [GAM+14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. *Generative Adversarial Nets*. 2014.
- [DR17] NT. Deshpande, S. Ravishankar. *Face Detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN*. 2017.
- [KRR+10] Ramesha K., K. B. Raja, Venugopal K. R., L. M. Patnaik. Feature Extraction based Face Recognition, Gender and Age Classification. 2010.
- [FHR+21] Stanislav Frolovab, Tobias Hinzcd, Federico Raueb, Jörn Heesb, Andreas Dengelab. *Adversarial text-to-image synthesis: A review*. 2021.
- [EAD+20] Hajar Emami, Majid Moradi Aliabadi, Ming Dong and Ratna Babu Chinnam. *SPA-GAN: Spatial Attention GAN for Image-to-Image Translation*. 2020.
- [KLA+20] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila. *Analyzing and Improving the Image Quality of StyleGAN*. 2020.
- [WES19] Mika Westerlund. *The Emergence of Deepfake Technology: A Review*. 2019.
- [ZGF+17] Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin. *Adversarial Feature Matching for Text Generation*. 2017.
- [DS15] A. Dosovitskiy, J. Tobias Springenberg. *Learning to generate chairs with convolutional neural networks*. 2015.
- [GDG+15] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, Daan Wierstra. *DRAW: A Recurrent Neural Network For Image Generation*. 2015.
- [SZS+20] Tianyang Shi, Zhengxia Zou, Xinhui Song, Zheng SongChangjian Gu, Changjie Fan, Yi Yuan. *Neutral Face Game Character Auto-Creation via*

- PokerFace-GAN*. 2020.
- [KST+22] Amina Kammoun, Rim Slama, Heidi Tabia, Tarek Ouni, Mohamed Abid. *Generative Adversarial Networks for face generation: A survey*. 2022.
- [KLA19] Tero Karras, Samuli Laine, Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Network*. 2019.
- [SGT+20] Yujun Shen, Jinjin Gu, Xiaoou Tang, Bolei Zhou. *Interpreting the Latent Space of GANs for Semantic Face Editing*. 2020.
- [VZR+20] Sushma Venkatesh , Haoyu Zhang , Raghavendra Ramachandra , Kiran Raja , Naser Damer , Christoph Busch. *Can GAN Generated Morphs Threaten Face Recognition Systems Equally as Landmark Based Morphs? - Vulnerability and Detection*. 2020.
- [ETI21] Hubert Etienne. *The future of online trust (and why Deepfake is advancing it)*. 2021.
- [YK21] Ufuk Yalçın, Duygu Akşit Karaçam. *Authorship in Deepfake Movies*. 2021.
- [LMH22] Vejay Lalla, Adine Mitrani and Zach Harned. *Artificial Intelligence: Deepfakes in the Entertainment Industry*. 2022.
- [GOO22] Google. *Overview of GAN Structure*. Prieiga internete: https://developers.google.com/machine-learning/gan/gan_structure. Žr. 2022-06-18
- [KAL+21] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, Timo Aila. *Alias-Free Generative Adversarial Networks (StyleGAN3)*. 2021
- [KAL+18] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen. *Progressive growing of GANs for improved quality, stability, and variation*. 2018
- [BDS19] Andrew Brock, Jeff Donahue, Karen Simonyan. *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. 2019.
- [MQZ+21] Liu Mingcong, Li Qiang, Qin Zekui, Zhang Guoxin, Wan Pengfei, Zheng Wen. *BlendGAN: Implicitly GAN Blending for Arbitrary Stylized Face*

- Generation*. 2021
- [WZL+18] Wayne Wu, Yunxuan Zhang, Cheng Li, Chen Qian, Chen Change Loy. *ReenactGAN: Learning to Reenact Faces via Boundary Transfer*. 2018.
- [NKH19] Yuval Nirkin, Yosi Keller, Tal Hassner. *FSGAN: Subject Agnostic Face Swapping and Reenactment*. 2019.
- [PGC+21] Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr. Dpfks, Carl Shift Facenheim, Luis RP, Jian Jiang, Sheng Zhang, Pingyu Wu, Bo Zhou, Weiming Zhang. *DeepFaceLab: Integrated, flexible and extensible face-swapping framework*. 2021.
- [CHW+21] Meng Cao, Haozhi Huang, Hao Wang, Xuan Wang, Li Shen, Sheng Wang, Linchao Bao, Zhifeng Li, Jiebo Luo. *UniFaceGAN: A Unified Framework for Temporally Consistent Facial Video Editing*. 2021.
- [AZM+22] Rameen Abdal, Peihao Zhu, Niloy J. Mitra, Peter Wonka. *Video2StyleGAN: Disentangling Local and Global Variations in a Video*. 2022.
- [SGZ+16] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen. *Improved Techniques for Training GANs*. 2016.
- [NKS21] Eric J. Nunn, Pejman Khadivi, Shadrokh Samvi. *Compound Fréchet Inception Distance for Quality*. 2021.
- [HRU+17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter. *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. 2017.
- [KSP22] Minguk Kang, Joonghyuk Shin, Jaesik Park. *StudioGAN: A Taxonomy and Benchmark of GANs for Image Synthesis*. 2022.
- [NM20] Jim Nilsson, Tomas Akenine-Möller. *Understanding SSIM*. 2020.

Priedai

1 priedas. IEEE Access pateiktas straipsnis

Eglynas, Tomas; **Lizdenis, Dovydas**; Raudys, Aistis; Jakovlev, Sergej; Voznak, Miroslav. Exploring Generative Adversarial Networks: Comparative Analysis of Facial Image Synthesis and the Extension of Creative Capacities in Artificial Intelligence // IEEE Access. Piscataway : IEEE. [Science Citation Index Expanded (Web of Science); Scopus; IEEE Xplore] [IF: **3,476**; AIF: 3,996; **kvartilis Q2** (2021, InCites JCR SCIE)] [CiteScore: 6,70; SNIP: 1,315; SJR: 0,927; kvartilis Q1 (2021, Scopus Sources)] [M.kr.: T 007]

[Bendrautoriaus indėlis.: 0,300]

For consideration in IEEE Access

Page 1 of 12



IEEE Access



Regular Manuscript

Exploring Generative Adversarial Networks: Comparative Analysis of Facial Image Synthesis and the Extension of Creative Capacities in Artificial Intelligence

Submission ID 4ba515b6-9512-4b30-a939-da7b44d20a3d
Submission Version Initial Submission
PDF Generation 23 Apr 2024 07:41:26 EST by Atypon ReX

Authors

Dr. Tomas Eglynas

Affiliations

• Marine Research Institute, Klaipėda University, 84 Herkaus Manto str., 92294 Klaipėda, Lithuania

Mr. Dovydas Lizdenis

Affiliations

• Marine Research Institute, Klaipėda University, 84 Herkaus Manto str., 92294 Klaipėda, Lithuania
• Faculty of Mathematics and Informatics, Vilnius University, Naugarduko str. 24, LT-03225, Vilnius, Lithuania

Dr. Aistis Raudys

Affiliations

• Faculty of Mathematics and Informatics, Vilnius University, Naugarduko str. 24, LT-03225, Vilnius, Lithuania

Dr. Sergej Jakovlev
Corresponding Author
Submitting Author

Affiliations

• Marine Research Institute, Klaipėda University, 84 Herkaus Manto str., 92294 Klaipėda, Lithuania
• Telecommunications Department, VSB-Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic

Exploring Generative Adversarial Networks: Comparative Analysis of Facial Image Synthesis and the Extension of Creative Capacities in Artificial Intelligence

Tomas Eglynas¹, Dovydas Lizdenis^{1,2}, Aistis Raudys², Sergej Jakovlev^{1,3}, and Miroslav Voznak^{1,3}

¹Marine Research Institute, Klaipėda University, 84 Herkaus Manto str., 92294 Klaipėda, Lithuania
²Faculty of Mathematics and Informatics, Vilnius University, Naugarduko str. 24, LT-03225, Vilnius, Lithuania
³Telecommunications Department, VSB-Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
Corresponding author: Sergej Jakovlev (e-mail: sergej.jakovlev@tu.ohb.cz)

ABSTRACT Neural networks represent a pivotal technological advancement in contemporary times, finding applications across diverse fields such as medicine, law enforcement, and information technology. These networks are designed to train algorithms using specified datasets to execute tasks autonomously, without the need for explicitly programmed instructions. A notable challenge in computational fields is emulating the distinctly human capacity for creativity—specifically, the ability to conceive and materialize novel and tangible entities. Humans possess the unique capability to imagine intricate worlds, environments, and characters, and to manifest these visions in various artistic forms, such as literature, music, and visual arts. To approximate this creative aptitude through artificial intelligence, sophisticated algorithms have been developed, among which generative adversarial networks (GANs) are prominent. GANs are particularly renowned for their ability to synthesize realistic human facial images. Despite their widespread use, there exists a paucity of comprehensive comparative studies on the current methodologies and outputs of face-generating GANs. This paper aims to elucidate the application spectrum of facial generation, delineate the operational principles of the original GAN framework, and survey existing variants tailored for facial synthesis. Furthermore, this study evaluated the performance of these models and analyzed the fidelity of the images they produce, thereby contributing to the broader understanding of generative capabilities in artificial intelligence.

INDEX TERMS Human image synthesis, Image processing, Computer graphics, Visualization, Photorealism

I. INTRODUCTION

Artificial neural networks (ANNs) have emerged as one of the most influential technologies of our era, extensively applied across various sectors including medicine [1], [2], [3], law enforcement [4], and information technology [5]. The fundamental principle of ANNs involves training algorithms with provided data to autonomously perform specific tasks. These networks excel in accurately recognizing objects and individuals in images and videos [6], [7], swiftly and precisely classifying photography [8], [9], [10], [11], identifying faces [8], [12], [13], and even determining characteristics such as gender or current

emotional state based on facial recognition [14]. Moreover, ANNs can predict future changes based on historical data [4], [15], [16], demonstrating their versatile applicability across diverse domains: ranging from various engineering solutions – to social domains [17]. Practically, neural networks can be adapted wherever it is feasible to train a computer to execute high computational complexity tasks [18]. However, one of the distinct human traits is the ability to create something new and tangible. Humans can envision diverse worlds, environments, people, and scenes, and translate these into books, music, and paintings, thereby

For consideration in IEEE Access

Page 2 of 12

Prof. Miroslav Voznak

Affiliations

• Marine Research Institute, Klaipėda University, 84 Herkaus Manto str., 92294 Klaipėda, Lithuania
• Telecommunications Department, VSB-Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic

Additional Information

Keywords

Computer graphics
Computer graphics

Human image synthesis
Human image synthesis

Image processing
Image processing

Subject Category

Computational and artificial intelligence
Computational and artificial intelligence

Computers and information processing
Computers and information processing

Imaging
Imaging

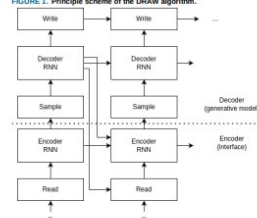
Files for peer review

All files submitted by the author for peer review are listed below. Files that could not be converted to PDF are indicated; reviewers are able to access them online.

Name	Type of File	Size	Page
IEEEAccess-publikacija_Dovydo_2023_v07.pdf	Main Document - PDF	621.4 KB	Page 3

algorithms such as Convolutional Neural Networks (CNNs) [19], [20], Deep Recurrent Attentive Writers (DRAW) [21] based upon Recurrent Neural Networks (RNNs) [22], [23], [24], [25], and Generative Adversarial Networks (GANs) [26], [27], [28], [29], [30], [31] are employed. Beginning with the CNN algorithm, it is possible to generate an image; an example explored is the generation of simple chair images. Initially, the neural network is trained using chair images as data. Upon training, by inputting specific parameters like chair type, representation, and color, the neural network can generate an image of a chair, ranging in size from 48x48 pixels to 128x128 pixels—though larger images can be produced at the expense of additional computational resources. The core concept of the DRAW algorithm involves using an encoder and decoder, which are recurrent networks, to sequentially read the input in parts considering previously completed steps and encode and store the obtained information. Subsequently, the input is decoded and reconstructed anew, much like an autoencoder [21]. Additionally, while reading the data, the encoder network is aware of the previously output decoder result, which informs it of the data that needs to be read next. Following the data reading phase, the DRAW algorithm can generate images from the stored and encoded data, creating new visuals or objects. The schematic of the DRAW algorithm is depicted in Figure 1.

FIGURE 1. Principle scheme of the DRAW algorithm.



In Figure 1, the schematic of the DRAW algorithm is depicted, illustrating all the steps previously described. Below the dashed line, the scanning and encoding steps are shown, which lead to the "sample" step where information is stored and from which images are subsequently

Additionally, arrows in the image indicate how the networks share information after each scanning phase. The Generative Adversarial Network (GAN), is currently one of the most popular for image synthesis [32]. Image synthesis plays a critical role and can be applied in fields such as art generation, computer design, photo editing, and virtual reality. GANs are also adept at creating high-quality images of human faces [33] and can be used to generate video clips featuring the faces of various individuals [34], transform from one photo to another, produce realistic text, enhance the resolution in images, convert text to image, and complete or augment parts of an image (known as image inpainting) [26], [35]. The general principle of GAN is illustrated in Figure 2.

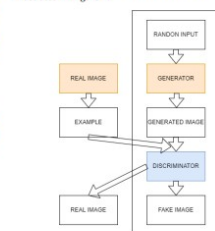


FIGURE 2. Principle scheme of the GAN algorithm.

This diagram shows the GAN architecture, which includes two competing networks: a generator and a discriminator. The generator receives random input to produce an image, which is then assessed by the discriminator along with a genuine image. The discriminator determines whether the generated image appears real. If deemed real, the image is presented as the result. If not, the generator learns from this feedback and attempts to create a new image that can convince the discriminator of its authenticity. This work will examine various GAN algorithms that generate faces, addressing the current lack of comprehensive comparative studies on facial generation algorithms. Furthermore, as previously mentioned, unlike DRAW or CNN, GANs can replicate the human ability to create new and realistic entities [36]. This study will first explore where facial generation is applied and review the GANs currently developed for this purpose. The operational principles of the first GAN proposed by Ian

This illustration depicts the architecture of traditional GAN and StyleGAN generators. In the traditional generator architecture, an initial parameter is included from which an image is generated. In the StyleGAN architecture, three new components are added: the first component is a mapping network where several initial vectors are introduced and transformed into new vectors w . These vectors are then randomly applied to a new normalization block, AdaIN, which represents the second addition. The third addition involves the implementation of random noise not only at the beginning but also at every subsequent step with varying values, allowing control over the 'style' of the generated image or face [33].

Further modifications in StyleGAN2 primarily altered the AdaIN function. It can be observed that this function now splits into two parts. In one part, the vector is normalized and introduced in style A , while in the other, weights are distributed according to style settings and then passed into a Conv 3x3 block, like the original StyleGAN architecture [37]. These changes in the generator allow for the creation of various faces stylized according to specific needs. However, this architecture is not perfect, as the generated images may contain certain artifacts that detract from their realism. Therefore, an updated version, StyleGAN2, introduces changes in the architecture aimed at producing higher-resolution images with fewer distortions. As demonstrated, this is achieved by modifying the AdaIN normalization block and the distribution of weights according to predefined style parameters.

Thus, StyleGAN provides the capability to generate realistic faces from given parameters with specific styles, though not all images are of high quality. Some generated images may contain artifacts or other unfitted details, leading to the introduction of the new StyleGAN2. This model addresses the discussed issues by minimizing the occurrence of artifacts in the generated images.

C. BigGAN ANALYSIS

The general GAN most significantly benefits from scaling up image dimensions. By doubling or even quadrupling the parameters and increasing the batch size eightfold during the training phase, higher-quality results are achieved.

Additionally, BigGAN modifies the architecture by enhancing scaling, improving conditioning, and boosting performance. With these modifications, the BigGAN model can generate realistic images, faces, and other objects, as it was trained with a large dataset featuring over 1,000 categories [38]. Despite being trained with extensive data, the BigGAN model has its shortcomings: training the model requires substantial resources, and the initial model was trained using over a hundred GPUs, indicating that training such a model with personal resources would minimally require four GPUs (G4iHub). The initially trained model was later made publicly available. Another issue with BigGAN is that the generated

images may include a significant number of flawed ones with artifacts or other discrepancies, as can be seen in Figure 6.

Thus, the BigGAN model, trained with a vast array of image categories, can produce high-quality and highly realistic images. However, such a large model might generate many flawed, irrelevant, and incomprehensible images while creating the desired photo.

D. PG-GAN (ProGAN) ANALYSIS

Progressive Growing of GANs (PG-GAN), also known as ProGAN, revolutionized the training process to make it more stable. A more stable training process results in higher-quality generated images. PG-GAN achieves this stability by starting the training with very low resolutions, such as 4x4, and gradually adding layers to increase the resolution of the output from the generator and the input image size for the discriminator [39].

This process continues until the desired image size is achieved. This method has proven very effective in creating high-quality synthetic images that appear realistic. Essentially, PG-GAN introduced three key enhancements:

- **Progressive Growth in Training:** The resolution is progressively increased during training: starting at a low resolution such as 4x4 and incrementally increasing to the desired size, such as 1024x1024.
- **Normalization Function Change to PixelNorm:** This modification helps in normalizing pixel values across the generated images to ensure consistency in quality.
- **Additional Minibatch Functionality in the Discriminator:** This feature improves the discriminator's ability to manage variations within a minibatch, enhancing its accuracy in distinguishing real from fake images.

As mentioned, the first point involves step-by-step modifications during the training—starting from a low resolution and gradually increasing the resolution with each output until reaching the desired image size, such as 1024x1024. The diagram of this process is seen in Figure 4.

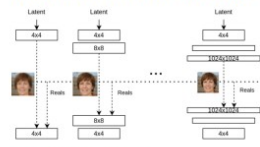


FIGURE 4. PG-GAN Training Framework.

TABLE I
COMPARISON OF GANs GENERATING FACES BASED ON INCEPTION SCORE (IS) AND FRÉCHET INCEPTION DISTANCE (FID)

GANs	Evaluation methods (Results Obtained from Other Sources)		Evaluation methods (Results Obtained from Facial Commentators)	
	IS	FID	IS	FID
StyleGAN2	5.17	7.70	7.24	31.68
BigGAN	166.5	7.4	5.97	212.77
PG-GAN	8.8	7.3	7.38	186.79

These results include ratings derived from both published articles where these models were originally introduced and from newly generated images using the same generative models. Upon reviewing the data, it is evident that there are discrepancies between the evaluations reported in the literature and those obtained from our experiments.

The IS, which quantifies the diversity of generated images, was highest for BigGAN in published results, reflecting its ability to produce a wide variety of image categories. However, our experiments showed that while BigGAN excels in diversity, it tends to generate fewer categories specifically related to human features, which impacts its IS rating. StyleGAN2, on the other hand, was found to generate the most realistic images according to the FID metric, indicating fewer discrepancies between generated and actual human images.

The observed differences in results could be due to several factors. For instance, in the case of StyleGAN2, a better IS score might suggest that a larger set of generated images was used in the study, potentially including a higher number of duplicates, which artificially inflates the diversity measure. Conversely, a higher FID in published studies may result from using a larger sample of images, which increases the likelihood of capturing realistic variations.

For BigGAN, the significantly lower IS and FID scores in our tests compared to those reported may be attributed to its application to a narrower range of categories, primarily focused on human-related images, unlike its typical use across thousands of varied categories. This specialization might not fully utilize BigGAN's capacity for generating diverse image types, thus reflecting poorer performance in our specific test setup. PG-GAN showed similar IS values to those reported, but significant differences in FID were noted, possibly due to the presence of minor artifacts in the generated images. While these artifacts are not overly prominent, they can affect the realism of the images and thus the FID score.

Summarizing the results from both the literature and our findings, BigGAN exhibits the highest diversity among the models tested. However, StyleGAN2 consistently produces the most realistic images, followed closely by PG-GAN and then BigGAN. Moving forward, the generated outcomes illustrate that StyleGAN2 not only produces a broader variety of images compared to other GANs but also maintains superior realism. The upcoming sections will present images generated by the BigGAN model, shown in Figure 5, further illustrating these differences.



FIGURE 5. Results Obtained from BigGAN.

BigGAN was used to generate faces within several broad categories that include human elements, such as tennis players, brides, glasses, and wigs because it lacks a specific category for faces. The results display certain resemblances to these generated categories, but they do not achieve realism in terms of facial generation. Following this, the results generated by PG-GAN are presented in Figure 6.



FIGURE 6. Results Obtained from PG-GAN.

PG-GAN's results appear more realistic than those generated by BigGAN, yet they also have their shortcomings, such as inaccurately generated hair artifacts within the images, and a significant number of duplicates. Next, the results generated by StyleGAN2 are presented in Figure 7.



FIGURE 7. Results Obtained from StyleGAN2.

The results obtained with StyleGAN2 are the most realistic compared to those generated by BigGAN and PG-GAN. Additionally, StyleGAN2 exhibits greater diversity and fewer artifacts in the images. While not all images are perfect, they present a wider variety of realistic faces without extensive searching among generated images. Considering these results, StyleGAN2 emerges as the best GAN for facial generation due to its ability to produce the most realistic images and the greatest diversity. This is also reflected in the obtained

The image in Figure 4 demonstrates a progressive training process where both the generator (G) and the discriminator (D) begin with low resolution. As training progresses, D and G simultaneously increase the resolution of the images. According to the authors, this process can reduce training time by 2 to 5 times, depending on the desired resolution of the outcome. Additionally, this method of training yields higher-quality images.

The next modification is the implementation of PixelNorm normalization during training to ensure that the training levels of the generator and discriminator do not diverge significantly. Otherwise, either the generator or the discriminator might learn more rapidly than the other and dominate, thus halting the training process and leading to unsatisfactory results.

Another change is applied to the discriminator—adding what the authors call a Minibatch feature, which allows the discriminator to remember previously generated images after each iteration. This helps the discriminator determine whether the generator's image is real or fake. It forces the generator to produce more diverse images in an attempt to deceive the discriminator. This method addresses one of the common problems with GANs, where trained models produce images with limited variety.

Thus, PG-GAN introduced modifications that help stabilize the training by adding the PixelNorm normalization block, preventing either the generator or discriminator from advancing too quickly and disrupting the training. Incremental resolution enhancement up to the predefined level also supports smoother training, helping both models progressively learn. Additionally, by incorporating the Minibatch feature in the discriminator, more varied and numerous results are achievable. All these changes enable the generation of realistic and high-quality images with greater diversity than typically seen.

StyleGAN offers the capability to generate realistic faces and control their styles using specific parameters. BigGAN was trained with a large volume of photos across numerous categories. This GAN enhanced overall training parameters and photo resolution, achieving high-quality, realistic images. However, a drawback of these BigGAN modifications is that training requires substantial resources. PG-GAN introduces a novel training approach that begins at low resolutions and incrementally increases to the desired resolution. It also introduced a new normalization function that enhances the stability of the model training and added a Minibatch feature in the discriminator, which forces the generator to produce a more diverse array of images.

E. GAN EVALUATION METRICS: INCEPTION SCORE (IS), FRÉCHET INCEPTION DISTANCE (FID)

Inception Score (IS) [10], [36], [40], Fréchet Inception Distance (FID) [10], [26], [32], [36], [40], [41], [42] evaluation metrics were chosen because they are among the primary objective methods of comparison used by authors in

scholarly articles. Additionally, these methods provide quantitative results that demonstrate the capabilities of a GAN model. The evaluation begins with the Inception Score (IS), which measures two main criteria:

- **Photo Quality:** This assesses whether the image reflects the type of the preceding images; for example, if one image contains a cat, subsequent images should also feature cats.
- **Photo Diversity:** This assesses whether the generated images are varied despite being of the same type, such as generating images of cats where each generated image features a different breed.

If the images produced by a GAN model satisfy both criteria, the IS result will be high. Theoretically, there is no maximum value for IS. If the generated images do not meet these criteria, the IS result will be low, with the minimum IS value potentially being 0. Therefore, the higher the IS result, the more diversity the GAN model can generate, capable of producing recognizable and general category images such as cats, dogs, faces, or other visuals. The overall result is determined based on the distribution of images. If we have generated an image from which it is determined what is seen and assigned to a category, for example, a dog. If most classified images were of dogs, that means that the generated images are primarily recognized as dogs, which would indicate a higher IS result because it signifies that the generated images are of high quality and recognizable as belonging to the same class. If the graph were evenly distributed, it would indicate a lower IS result because the images are not of good quality, thus not understanding the general generated category. The limitation of IS is that it evaluates overall image quality and diversity but does not assess realism.

The realism score is provided by the Fréchet Inception Distance (FID) evaluation method. FID assesses a collection of generated images against a collection of real images from the generated domain. This evaluation, combined with the IS assessment, helps determine whether a GAN model produces high-quality, varied, and realistic images. FID is calculated by comparing the distances between the vectors of the generated images and the real images according to the generated category, thus assessing realism. The lower the FID score, the more realistic the generated images; a higher FID score indicates that the generated image is less realistic.

III. RESULTS

In conclusion, StyleGAN2, BigGAN, and PG-GAN for facial generation are detailed using the Inception Score (IS) and Fréchet Inception Distance (FID) metrics, with the results displayed in Table 1. Here the performance of various GAN models using Inception Score (IS) and Fréchet Inception Distance (FID) metrics is compared.

Inception Score (IS) and Fréchet Inception Distance (FID) metrics. PG-GAN and BigGAN follow, with BigGAN being less suitable for facial generation as its images lack realistic outcomes. PG-GAN, while capable of generating faces that resemble humans, often produces similar and artifact-ridden results compared to StyleGAN2. Moreover, StyleGAN2's ability to control facial styles—such as general features, age, and skin tone—offers a significant advantage for generating specific faces or datasets more easily and flexibly.

IV. CONCLUSIONS

In this part of the study, we explored image evaluation metrics such as the Inception Score (IS) and the Fréchet Inception Distance (FID), which are used to assess the diversity, quality, and realism of generated images. By employing these evaluation methods, we compared the images produced by StyleGAN2, BigGAN, and PG-GAN models. The results indicated that the StyleGAN2 generative model achieved the highest ratings, although BigGAN generated the greatest diversity of images. In terms of realism, as indicated by the results, StyleGAN2 also produced the most lifelike images.

Reviewing these generative models visually, it was evident that StyleGAN2 consistently delivered superior outcomes, as exemplified in Figure 7. While these results are impressive, a closer analysis of the generated images reveals even more visually appealing faces, examples of which are presented in Figure 8.



FIGURE 8. Realistically Generated Faces Selected from StyleGAN2.

As demonstrated in Figure 8, the faces depicted are generated with remarkable realism, featuring only minimal noticeable artifacts, creating the impression that the images could be of real people. However, despite the presence of highly realistic images, there are instances where generated artifacts are visible, detracting from the realism of the photographs, as illustrated in Figure 9.



FIGURE 9. Non-Realistically Generated Faces Selected from StyleGAN2.

Based on Figure 9, it is apparent that the depicted individuals are artificially generated, as evidenced by blurred facial details or features that do not naturally occur. In addition to the facial features, the hair structure and background also indicate that the images are synthesized. This suggests that while the images appear artificial, they provide useful insights into the capabilities and limitations of generative models. In summary, StyleGAN2 proves to be effective for facial generation; the results obtained from this model are generally realistic, although, like other models, it occasionally produces errors during generation. These errors can manifest as unnatural features or distorted elements, underscoring the ongoing challenges in achieving flawless realism in generated images.

In conclusion, this study has demonstrated the efficacy of generative adversarial networks (GANs) in producing facial images, with a particular focus on the performance of StyleGAN2, BigGAN, and PG-GAN. The evaluation using metrics such as the Inception Score and the Fréchet Inception Distance revealed that StyleGAN2 consistently outperformed the other models in terms of image quality and realism, albeit with some occurrence of artifacts. While BigGAN excelled in generating a diverse array of images, it occasionally produced images with noticeable distortions. PG-GAN was notable for its high-quality image generation, modifying the training principles of traditional GANs to achieve better results.

Despite the advancements in facial generation technology demonstrated by these models, the presence of artifacts and occasional unrealistic features highlights the challenges that still lie ahead in the field of synthetic image generation. Future research should focus on refining these models to minimize errors and enhance the realism and utility of generated images for practical applications in various domains such as digital media, entertainment, and security.

REFERENCES

- [1] W. Xu, Y. Li, F. H. Xu, and K. K. L. Wong, "Medical image fusion using enhanced cross-visual cortex model based on artificial selection and impulse-coupled neural network," *Comput Methods Programs Biomed.*, vol. 229, p. 107364, Feb. 2021, doi: 10.1016/j.cmpb.2021.107364.
- [2] S. Yantong, Y. Shilin, Z. Xiangling, and Y. Jinhua, "Edge Detection Algorithm of MRI Medical Image Based on Artificial Neural Network," *Procedia Comput. Sci.*, vol. 208, pp. 136–144, Jan. 2022, doi: 10.1016/j.procs.2022.101021.
- [3] S. Abee, H. Okae, and K. J. Kallala, "Paradigm shift from Artificial Neural Networks (ANNs) to deep Convolutional Neural Networks (DCNNs) in the field of medical image processing," *Expert Syst Appl.*, vol. 244, p. 122863, Jun. 2024, doi: 10.1016/j.eswa.2023.122863.
- [4] J. Chen, W. Tao, Z. Jiao, P. Wang, and Y. Jia, "Traffic accident duration prediction using multi-mode data and ensemble deep learning," *Inforsys*, vol. 10, no. 4, p. e25657, Feb. 2024, doi: 10.1016/j.inforsys.2024.e25657.
- [5] L. Li, X. Shang, B. Du, Y. Wang, and B. Ran, "A deep fusion model based on restricted Boltzmann machines for traffic accident duration prediction," *Eng Appl Artif Intell*, vol. 93, Aug. 2020, doi: 10.1016/j.engappai.2020.103686.

[6] K. Oh et al., "Artificial intelligence - Based video traffic policing for next generation networks," *Simul Model Pract Theory*, vol. 121, p. 102650, Dec. 2022, doi: 10.1016/j.smp.2022.102650.

[7] W. He et al., "A state-of-the-art survey of artificial neural networks for Whole-slide Image analysis: From popular Convolutional Neural Networks to potential visual transformers," *Comput Biol Med*, vol. 161, p. 107034, Jul. 2023, doi: 10.1016/j.cmb.2023.107034.

[8] C. Goutan and K. R. Seta, "Facial emotion recognition using Handcrafted features and CNN," *Procedia Comput Sci*, vol. 218, pp. 1295-1303, Jan. 2023, doi: 10.1016/j.procs.2023.01.188.

[9] R. Kumar, J. Sotelo, K. Kumar, A. de Brébisson, and Y. Bengio, "ObamaNet: Photo-realistic lip-syn from text," Dec. 2017.

[10] M. Tashbakhsh, W. Lee, and S. Lee, "Talking human face generation: A survey," *Expert Syst Appl*, vol. 219, p. 119678, Jun. 2023, doi: 10.1016/j.eswa.2023.119678.

[11] S. Kaur et al., "Face mask recognition system using CNN model," *Neuroscience Informatics*, vol. 2, no. 3, p. 100035, Sep. 2022, doi: 10.1016/j.neuri.2021.100035.

[12] S. Mianaei, M. Mianaei, and A. Abdolrahimi, "Deep-emotion: Facial expression recognition using attentional convolutional networks," *Sensors*, vol. 21, no. 9, May 2021, doi: 10.3390/S21090046.

[13] S. A. Prone, N. A. Ragnav, M. R. Islam, D. Astvithan, and A. J. Jagatheesan, "Emotion detection using machine learning (ML) and deep learning (DL) techniques: A systematic review," *Natural Language Processing Journal*, vol. 6, p. 100057, Mar. 2024, doi: 10.1016/j.nlp.2024.100057.

[14] A. P. Fard and M. H. Mahoor, "Ad-Center Adaptive Contrast-Based Loss for Facial Expression Recognition in the Wild," *IEEE Access*, vol. 10, pp. 26756-26768, 2022, doi: 10.1109/ACCESS.2022.3156598.

[15] F. He, Q. Yang, J. Yang, Z. Liu, J. Shao, and G. Wang, "Incorporating multiple grid-based data in CNN-LSTM hybrid model for daily runoff prediction in the source region of the Yellow River Basin," *J. Hydrol Res Stud*, vol. 51, p. 101652, Feb. 2024, doi: 10.1016/j.jhrs.2023.101652.

[16] M. Scholz, B. Boughatta, and F. Wendel, "DetEker: Mask R-CNN based neural network for energy plant identification on aerial photographs," *Energy and AI*, vol. 5, p. 100069, Sep. 2021, doi: 10.1016/j.egyai.2021.100069.

[17] S. Bhattarai and R. Bhandaryadhyay, "Modelling crash severity outcomes for low speed urban roads using back propagation - Artificial neural network (BP - ANN): A case study in Indian context," *IATSS Research*, vol. 47, no. 3, pp. 362-400, Oct. 2023, doi: 10.1016/j.iatssr.2023.08.002.

[18] E. Sanna, R. Tognolini, and G. Iaccarino, "Uncertainty quantification in autoregressive predictions: Applications in aerodynamics," *J. Comput Phys*, vol. 406, p. 112951, Jun. 2024, doi: 10.1016/j.jcp.2024.112951.

[19] Z. Li et al., "Extended Abstract of SCNN: A semantic CNN parser for code comment generation," 2023, doi: 10.1109/water673.2023.00099.

[20] Z. Li et al., "SCNN: A semantic CNN parser for code comment generation," *Journal of Systems and Software*, vol. 181, 2021, doi: 10.1016/j.jss.2021.111036.

[21] K. Ganga, J. Dambhala, A. Graves, D. J. Renneke, and D. Wierstra, "DRAW: A Recurrent Neural Network For Image Generation," Feb. 2015, [Online]. Available: <http://arxiv.org/abs/1502.04623>.

[22] A. Ouyens et al., "Restructuring Recurrent Neural Networks for Long Sequences," Mar. 2023, [Online]. Available: <http://arxiv.org/abs/2301.06349>.

[23] A. Kumar, N. Gaur, S. Chakraverty, M. H. Ashraf, P. Uthasankar, and M. Uthasankar, "Analysis of spectrum sensing using deep learning algorithms CNN and RNN," *Int J Smart Engineering Systems*, vol. 15, no. 3, Mar. 2024, doi: 10.1016/j.ijses.2023.102505.

[24] S. Yadav, H. Hashmi, D. Vekariya, Z. A. K. N., and V. F. J., "Mitigation of attacks via improved network security in IoT network environment using RNN," *Measurement: Sensors*, vol. 12, p. 101046, Apr. 2024, doi: 10.1016/j.msens.2024.101046.

[25] E. Yauri-Llanusa, M. Castillo-Cara, I. Onozco-Barbosa, and R. Garcia-Castro, "Generative Adversarial Networks for text-to-face synthesis & generation: A quantitative - qualitative analysis of Natural Language Processing encoders for Spanish," *Inf Process Manag*, vol. 61, no. 3, p. 103667, May 2024, doi: 10.1016/j.ipm.2024.103667.

[26] N. Day, A. Chen, and S. Ghafarian, "GRASP: EQUIVARIANT GENERATIVE ADVERSARIAL NETWORKS," *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[27] L. Wang, Y. Li, J. Liu, J. Peng, Q. Zhang, and W. Fu, "Research on fault diagnosis of industrial robots based on generative adversarial network," *Physical Communication*, p. 102555, Apr. 2024, doi: 10.1016/j.phycom.2024.102555.

[28] L. Yan and C. Lin, "Matrix Wasserstein distance generative adversarial network with gradient penalty for fast low-carbon economic dispatch of novel power systems," *Energy*, p. 131357, Apr. 2024, doi: 10.1016/j.energy.2024.131357.

[29] J. He, Y. Xia, Y. Pan, and Y. Wang, "Adaptive weighted generative adversarial network with attention mechanism: A transfer data augmentation method for tool wear prediction," *Mech Syst Signal Process*, vol. 212, p. 111288, Apr. 2024, doi: 10.1016/j.ymssp.2024.111288.

[30] J. J. Goodfellow et al., "Generative Adversarial Networks," Apr. 2014, [Online]. Available: <http://arxiv.org/abs/1406.2651>.

[31] J. Harkness et al., "Deformation equivariant cross-modality image synthesis with paired non-aligned training data," *Med Image Anal*, vol. 90, p. 102948, Dec. 2023, doi: 10.1016/j.media.2023.102948.

[32] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 12, 2021, doi: 10.1109/TPAMI.2020.2970919.

[33] P. Li, H. Zhao, Q. Liu, P. Yang, and L. Zhang, "ToliteTalk: Multimodal-driven talking face video generation," *Computers and Electrical Engineering*, vol. 114, p. 109495, Mar. 2024, doi: 10.1016/j.compeleceng.2023.109495.

[34] J. Agnew, J. Herrera, H. Tao, and X. Zhu, "A survey and taxonomy of adversarial neural networks for text-to-image synthesis," *Inf Int J Interdiscip Res Data Min Knowl Discov*, vol. 10, no. 4, Jul. 2020, doi: 10.1002/WIDM.1345.

[35] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," *7th International Conference on Learning Representations, ICLR 2019*, 2019.

[36] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1912.04958>.

[37] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," Apr. 2018, [Online]. Available: <http://arxiv.org/abs/1809.11096>.

[38] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," Apr. 2017, [Online]. Available: <http://arxiv.org/abs/1710.10196>.

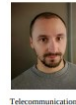
[39] N. Renkan, J. H. Giraldo, T. Brownman, S. Rodriguez-Barahona, E. Ramirez, and C. Izasa, "Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning," *Eng Appl Artif Intell*, vol. 124, p. 106635, Sep. 2023, doi: 10.1016/j.engappai.2023.106635.

[40] L. F. Barrett and C. E. Thomas, "Fréchet AutoLatent Distance: A new approach for evaluation of Generative Adversarial Networks," *Computer Vision and Image Understanding*, vol. 235, p. 103768, Oct. 2023, doi: 10.1016/j.cviu.2023.103768.

[41] E. J. Nune, P. Khadivi, and S. Samavi, "Compound Fréchet Inception Distance for Quality Assessment of GAN Created Images,"



TOMAS EGLYVAS is a native of Lithuania, completed his education at Klaipėda University, earning a Bachelor's degree in Informatics Engineering in 2010 and a Master's in Technical Information Systems Engineering in 2012. He received his Ph.D. in Transport Engineering through the Lithuanian Joint Transport Engineering Ph.D. program (KU, VGTU, ASU) in 2017. Presently, he holds the position of Senior Researcher at the Marine Research Institute of Klaipėda University and serves as CEO of Inozecha Ltd., a company based in Lithuania that develops intelligent industrial solutions. Dr. Eglyvas's research interests include intelligent control systems, Legation 4.0, control theory, electronics, and IoT applications in Industry 4.0. He has authored or co-authored numerous research articles published in SCOPUS and CA-indexed journals.



SERGEJ JAKOVLEV was born in Lithuania, completed his higher education at Klaipėda University, where he earned a Bachelor's degree in Informatics Engineering in 2009, followed by a Master's degree in Technical Information Systems Engineering in 2011. He later obtained his Ph.D. in Transport Engineering in 2016. Currently, he serves as a Chief Researcher at the Marine Research Institute of Klaipėda University and a Researcher at the Department of Telecommunications, VSB-Technical University of Ostrava. A prolific contributor to his field, Dr. Jakovlev is actively involved in several editorial boards and international science conferences, such as the International Science Conference on Computer Networks. He also reviews prestigious CA and SCOPUS journals including *Soft Computing* and *IEEE Transactions on Neural Networks and Learning Systems*. His research primarily focuses on intelligent transportation systems, AI theory and applications, Big Data analytics tools and methods, and wireless sensor networks. Dr. Jakovlev has authored or co-authored over thirty articles in SCOPUS and CA-indexed journals and proceedings.



DOVYDAS LIZDENIS was born in Lithuania and completed his education at Klaipėda State University of Applied Sciences in 2020 earning him a Professional Bachelor's Honors diploma in Informatics. Currently studying at Vilnius University getting a Master's degree in Informatics and serving as the CEO of Ltd., Digital Moon, a company based in Lithuania that develops software and web-based solutions. Dovydas' research interests are Software engineering, AI applications in engineering, and Computer Science domains.



MIROSLAV VOZNAK (IEEE Senior Member) is a professor in the Faculty of Electrical Engineering and Computer Science at the VSB - Technical University of Ostrava. He received his Ph.D. degree in Telecommunications in 2002, achieved habilitation in 2009, and was appointed full professor in 2017. His research interests include IoT, 5G/6G, wireless networks, network security, and big data analytics in networks. According to WoS, he has published more than 200 articles in SCIE journals such as *IEEE Communications Surveys & Tutorials*, *IEEE Journal on Selected Areas in Communications*, and *IEEE Communications Magazine*. He participated in seven projects within EU funding programs, mostly as an institutional coordinator, and more than twenty national projects. Since 2020, he has been ranked regularly among the World's Top 2% Scientists (Career Impact 20) by Stanford University in the subfield of Networking & Telecommunications.



AISTIS RAUDYS earned his PhD from Vilnius University in Lithuania, where his research focused on developing methods for extracting features from multivariate data. Currently, he serves as a professor at the Faculty of Mathematics and Informatics at Vilnius University, where he teaches courses related to algorithmic trading and robotics. In addition to his academic role, Aistis is the CEO and co-founder of AAI-Labs, an AI startup dedicated to the development of technologies in the fields of voice recognition, risk assessment, and transport optimization. Before his current position, Aistis gained experience as a researcher working with prominent financial institutions, including Inozecha Bank, Société Générale, and BNP Paribas, London, UK. Aistis Raudys' primary interests lie in financial engineering, robotics, automated trading, and artificial intelligence, and he has authored 50 publications in these areas of engineering, robotics, automated trading, and artificial intelligence.