

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Marijana Davitkova

**Zamenjava obrazov izbranih likov
slovenskega filma**

DIPLOMSKO DELO

VISOKOŠOLSKI STROKOVNI ŠTUDIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: pred. dr. Borut Batagelj

Ljubljana, 2023

COPYRIGHT. Rezultati diplomske naloge so intelektualna lastnina avtorja in matične fakultete Univerze v Ljubljani. Za objavo in koriščenje rezultatov diplomske naloge je potrebno pisno privoljenje avtorja, fakultete ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil L^AT_EX.

Kandidat: Marijana Davitkova

Naslov: Zamenjava obrazov izbranih likov slovenskega filma

Vrsta naloge: Diplomska naloga

Mentor: pred. dr. Borut Batagelj

Opis:

V zadnjem času je na voljo veliko metod in programske opreme, ki omogočajo zamenjavo obrazov in s tem ustvarjanja ponarejenih video posnetkov. Preučite različne metode za zamenjavo obrazov in izbrano metodo preizkusite na programu, ki bo izbranemu liku iz slovenskega filma zamenjal obraz z obrazom uporabnika.

Title: Changing the faces of selected Slovenian film characters

Iskreno se zahvaljujem mentorju pred. dr. Borutu Batagelju za mentorstvo, pomoč in brezkončno potrpežljivost. Posebna zahvala moji družini in prijateljem za njihovo neprecenljivo podporo in spodbudo.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Tehnologija globokih ponaredkov	3
2.1	Zgodovina tehnologije globokih ponaredkov	4
2.2	Kategorije globokih ponaredkov	5
3	Uporaba tehnologij globokih ponaredkov	9
3.1	Kinematografija	9
3.2	Jezikovne ovire	10
3.3	Izobrazba	11
3.4	Nevarnosti tehnologij globokih ponaredkov	12
4	Pregled metod	15
4.1	Generativna nasprotniška mreža	15
4.2	Variacijski samokodirnik	16
5	Pregled pristopov	19
5.1	Nekajkratno učenje	19
5.2	Omrežje X2Face	20
5.3	Model gibanja prvega reda	21
5.4	Metoda SimSwap	21

6 Pregled programske opreme	23
6.1 DeepFakes Faceswap	23
6.2 DeepFaceLive	25
6.3 Reface	25
7 Metoda SimSwap	27
7.1 Arhitektura, na kateri temelji metoda	27
7.2 Zamenjava poljubne identitete	28
7.3 Funkcija izgube	30
7.4 Rezultati	33
8 Izdelava programa za zamenjavo obraza	39
8.1 Ideja	39
8.2 Razlike med SimSwap in našim programom	40
8.3 Pripravljanje videa	41
8.4 Zamenjava obraza	43
8.5 Primer delovanja	44
8.6 Analiza uspešnosti	44
9 Zaključek	49
Literatura	51

Seznam uporabljenih kratic

kratica	angleško	slovensko
GAN	Generative Adversarial Network	Generativna nasprotniška mreža
VAE	Variational Auto-Encoders	Variacijski samokodirnik
VAE-GAN	Variational Auto-Encoders Generative Adversarial Network	Variacijska generativna nasprotniška mreža samokodirnika
FSL	Few Shots Learning	Nekajkratno učenje
CNN	Convolutional Neural Network	Konvolucijska nevron-ska mreža

Povzetek

Naslov: Zamenjava obrazov izbranih likov slovenskega filma

Avtor: Marijana Davitkova

Namen diplomske naloge je ustvariti program, ki bo znal obrazne poteze uporabnika hitro in prepričljivo zlepiti na obraze igralcev nekaterih znanih klasikov slovenske kinematografije. Po predhodnem učenju deluje program tudi z drugimi videoposnetki. Diplomska naloga uporablja algoritmom SimSwap za doseganje rezultatov in kot vhod vzame eno samo sliko uporabnika in ciljni videoposnetek ter ustvari globoko ponarejen videoposnetek, v katerem je obraz uporabnika namesto obraza enega od likov. Nastali videoposnetek in postopek sta zadovoljiva, z razmeroma hitrim postopkom in prepričljivim končnim ponarejenim videom v primerjavi s sorodnimi deli.

Ključne besede: globoki ponaredki, multimedia, manipulacija slik.

Abstract

Title: Changing the faces of selected Slovenian film characters

Author: Marijana Davitkova

The subject of this seminar is to create a program that will be able to quickly and convincingly paste the facial features of the user onto the faces of the actors of some well-known classics of Slovenian cinematography. This program works with other videos as well, but it needs to be trained first. The seminar uses the SimSwap algorithm to achieve the results and takes a single image of the user and the target video as input, using them to produce a deep faked video in which the user's face is on the place of one of the characters' faces. The resulting video and process was satisfactory, with a relatively fast running time, and convincing deep faked video, compared to contemporary works.

Keywords: deepfakes, multimedia, image manipulation.

Poglavlje 1

Uvod

V današnjem svetu gledanje in spremljanje filmov in serij ljudem vzame precej velik del prostega časa. Ne glede na to, ali občinstvo potrebuje smeh ali strah, nekaj drame ali morda akcijo v vsakdanjem življenju, je vsaka emocija v obliki filma ali serije oddaljena le en klik. Vendar pa je bilo še ne tako dolgo nazaj nekoliko drugače. Izbera filmov je bila majhna, za njihov ogled pa je bil potreben obisk kina ali pa veliko potrpljenja, da smo si film ogledali na domačem televizorju. Ta diplomska naloga je namenjena manipulaciji filmov, da lahko kot junak filma nastopi druga oseba.

Glavna funkcija aplikacije je zamenjava obraza enega od likov iz filma z obrazom uporabnika. Čeprav je tehnično mogoče aplikacijo uporabljati tudi s sodobnimi filmi in serijami, bo največji poudarek pri njeni uporabi na nekaterih brezčasnih klasikah iz preteklega stoletja.

Glavni motiv te diplomske naloge je želja po močnejši povezavi s kinematografijo preteklosti na način, ki je najbližji in najbolj domač novejšim generacijam – na virtualen način.

Izdelava te aplikacije bo sestavljena iz več metod in tehnik, kot so tehnologije globokega učenja in nevronske mreže, ki se bodo uporabljale v različnih fazah aplikacije, kot so zaznavanje obrazov, prepoznavanje obrazov in zamenjava obrazov.

Poglavlje 2

Tehnologija globokih ponaredkov

Fotografska manipulacija obstaja že od izuma fotoaparata. Ljudje so poskušali narediti groteskne, zanimive in nenavadne učinke z uporabo rekvizitov med fotografiranjem ali v procesu razvijanja poslikanih fotografij. Z napredovanjem tehnik za zajemanje slik so vzporedno rasle in napredovale tudi manipulativne metode. Najnovejša, najsodnejša tehnologija, ki se uporablja za manipulacijo slik in video posnetkov v današnjem času, je tehnologija globokih ponaredkov (angl. DeepFake). Sama beseda v angleščini je sestavljena iz dveh izrazov, Deep Learning, kar pomeni globoko učenje in Fake, kar pomeni, da posnetek ali slika, ki prihaja kot izdelek, ni izvirnik, ampak ponaredek.

Globoko učenje je vrsta strojnega učenja, ki uporablja več nivojev za učenje modela, pri čemer vsak nivo postopoma iz surovega vnosa izvleče funkcije višje ravni kot prejšnja. Metode globokega učenja podpirajo nadzorovano, polnadzorovano in nenadzorovano učenje ter temeljijo na umetnih nevronskih mrežah.

Globoki ponaredki so slike ali videoposnetki, ki so bili obdelani tako, da je obraz osebe prilepljen na videoposnetek druge osebe, zaradi česar se zdi, da oseba, katere obraz je bil nameščen, dejansko izvaja dejanja iz videoposnetka,

kamor je bila prilepljena.

2.1 Zgodovina tehnologije globokih ponaredkov

Računalniški vid kot podpodročje računalništva v veliki meri zajema akademiske raziskave, ki se ukvarjajo s tehnologijami globokih ponaredkov. Najzgodnejši projekt na to temo in prvi sistem, ki je tovrstno oživljanje obraza naredil povsem avtomatsko, je bil objavljen leta 1997 in se je imenoval Video Rewrite Program [6]. V tem projektu so uporabljene različne tehnike strojnega učenja, da bi se ustvarila povezava med obliko obraza osebe v videu in zvokom, ki ga je ustvarila na videu. S to tehniko so se spremajali obstoječi videoposnetki ljudi [5], ki se pogovarjajo, tako da se zdi, da oseba izgovarja besede z drugega zvočnega posnetka.

Tehnologija je od takrat zelo napredovala, sodobni projekti iz tega akademskega področja pa se osredotočajo na izboljšanje metod in tehnik za razvoj tehnologije globokih ponaredkov in tako ustvarjajo bolj realistične videe. Primer tega je program Face2Face [36], kjer se videoposnetek obraza subjekta spremeni tako, da ga prikazuje, kot da imitira obrazne poteze in izraze druge osebe v realnem času. Glavni prispevek tega projekta na akademskem področju je prva metoda za imitacijo obrazne mimike v realnem času in za to je projekt Face2Face uporabljal posebno kamero, ki zajema globino, te metode pa ne moremo replicirati z običajno kamero.

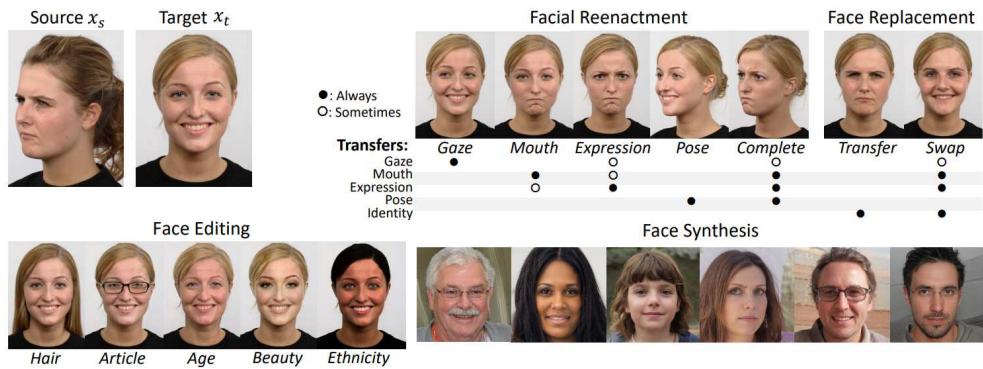
Drug pristop k problemu je bil narejen v programu Synthesizing Obama [34] iz leta 2017, ki uporablja video gradivo nekdajnega predsednika ZDA Baracka Obame in ga manipulira tako, da prikazuje, kako izgovarja besede, ki so izrečene v ločenem zvočnem posnetku [33]. Glavni prispevek na področju tehnologije globokih ponaredkov, kot je navedeno v tej nalogi, je njena realistična tehnika, ki se uporablja za sintetiziranje različnih oblik ust iz zvočnih posnetkov.

Naslednji korak v razvoju tehnologije globokih ponaredkov je bil nare-

jen leta 2018, ko so raziskovalci na kalifornijski univerzi Berkeley objavili članek [7] o programu AI, ki posnema plesne gibe iz videa, kjer se vidi celotno telo osebe. Avtorji ne uporabljajo izraza DeepFake in namesto tega tehnologijo označujejo kot "Do as I do" ("Naredi kot jaz"). To popelje tehnologije za obdelavo slik na naslednji korak, pri čemer se fokus razširi od samega obrazu subjekta na celotno telo.

2.2 Kategorije globokih ponaredkov

Na podlagi konteksta človeških vizualnih elementov so globoki ponaredki razdeljeni na kategorije: imitacija (angl. facial reenactment), zamenjava (angl. face replacement), urejanje (angl. face editing) in sinteza (angl. face synthesis) [23]. Primeri porazdelitve so vidni na sliki 2.1, kjer sta za imitacijo in zamenjavo obrazu uporabljeni izvorna slika (angl. source) x_s in ciljna slika (angl. target) x_t . Za urejanje obrazu je pa uporabljena samo ciljna slika x_t .



Slika 2.1: Primeri imitacije, zamenjave, urejanja in sinteze obrazov [23].

2.2.1 Imitacija

Imitacija se zgodi, ko se podoba izvorne identitete uporablja za upravljanja izraza, ust, telesa, poze ali pogleda podobe ciljne identitete. Poleg tega

jo lahko razdelimo na imitacijo izraza, imitacijo ust, imitacijo pogleda, imitacijo poze in celotno imitacijo. To lahko vidimo na sliki 2.1, pri "Facial Reenactment".

Imitacija izraza

Je najpogostejsa med oblikami imitacije. Predstavlja podobo izvorne identitete, ki je osnova za izražanje podobe ciljne identitete. Ker je ta tehnologija običajno kombinirana tudi z imitacijo ust in imitacijo poze, ponuja veliko prilagodljivost in jo pogosto najdemo v filmih in video igrah. Primer tega lahko vidimo na sliki 2.1, pri "Facial Reenactment" v kategoriji "Expression".

Imitacija ust

Medtem ko imitacija ust ni dobro znan izraz, je sinhronizacija zelo znan izraz in sinonim za imitacijo ust. Posnemanje je, ko usta subjekta iz ciljne slike presnamejo subjektova iz izvorne slike ali so prilagojena zvočnemu vhodu, ki vsebuje govor. To se večinoma uporablja pri prevajanju in montaži v filmih. To je vidno na sliki 2.1, pri "Facial Reenactment" v kategoriji "Mouth".

Imitacija pogleda

Ta vrsta imitacije se zgodi, ko oči in veke ciljnega subjekta temeljijo na očeh in vekah subjekta iz izvorne slike. Imitacija pogleda se običajno uporablja za izboljšanje fotografij ali samodejno ohranjanje očesnega stika med video intervjuji. To lahko vidimo na sliki 2.1 pri "Facial Reenactment" v kategoriji "Gaze".

Imitacija poze

V nasprotju s prej omenjenimi imitacijami, ki se ukvarjajo z obrazom, se imitacija poze ukvarja s položajem glave. Najde svojo uporabo pri izboljšanju prepoznavne ljudi iz varnostnih posnetkov, kjer imamo ljudi zajete iz različnih

pogledov. Primer tega lahko vidimo na sliki 2.1 pri ”Facial Reenactment” in ”Pose”.

Celotna imitacija

Nazadnje celotna imitacija obravnava celotno telo subjekta iz ciljnega posnetka, ki ga upravlja telo iz vira. Znana je tudi kot prenos poze ali sinteza človeške poze. To je vidno na sliki 2.1 pri ”Facial Reenactment” v kategoriji ”Complete”.

2.2.2 Zamenjava

Tehnologija zamenjave deluje na principu nadomestitve vsebine ciljne slike z vsebino izvorne slike, pri čemer se ohrani identiteta izvora. Lahko se pojavi v dveh oblikah: kot prenos ali vzajemna zamenjava. To lahko vidimo na sliki 2.1 pri ”Face Replacement”.

Prenos

Ta oblika zamenjave pomeni prenos identitete izvorne slike na osebo iz ciljne slike in se običajno uporablja v modni industriji za vizualizacijo posameznika v različnih oblačilih. Primer tega lahko vidimo na sliki 2.1 pri ”Face Replacement” v kategoriji ”Transfer”.

Vzajemna zamenjava

Vzajemna zamenjava je proces, v katerem ciljni video posnetek upravlja vsebino, ki jo nase prenese izvorna slika, tako da se poleg identitete prenese še izraz in druge značilnosti izvornega obraza. To je vidno na sliki 2.1 pri ”Facial Reenactment” v kategoriji ”Swap”.

2.2.3 Urejanje

Metode urejanja vključujejo dodajanje, odstranjevanje ali spreminjanje določenih lastnosti podobe subjekta iz ciljnega videa. Primer tega je spreminjanje

oblačil, barve oči, dolžina ali stil las in podobne podrobnosti. To lahko vidimo na sliki 2.1 pri "Face Editing".

2.2.4 Sinteza

Metoda sinteze se ukvarja z ustvarjanjem novih identitet in se ustvarja brez ciljnega subjekta. Ta metoda se večinoma uporablja za ustvarjanje neobstoječih slik obrazov. Primer tega lahko vidimo na sliki 2.1 pri "Face Synthesis".

Poglavlje 3

Uporaba tehnologij globokih ponaredkov

3.1 Kinematografija

Tehnologije globokih ponaredkov najdejo svojo uporabo predvsem v digitalnem svetu umetnosti. Uporabljene so bile v skoraj vseh hollywoodskih filmih, narejenih po devetdesetih let prejšnjega stoletja z namenom približati zabavo resnični realnosti. Najpogostejsa uporaba globokih ponaredkov je povezana s sinhronizacijo ustnic [12]: če je treba zvočni posnetek prizora znova posneti, igralci namesto snemanja celega prizora posnamejo samo zvok, nato pa se njihovi obrazi spremenijo z uporabo tehnologije globokih ponaredkov tako, da se zdi, kot da dejansko govorijo besedilo z zvočnega posnetka. 3D modeliranje je tudi področje, ki uporablja to tehnologijo, saj pomaga, da so značilnosti modela videti bolj realistične in verjetne. Pomaga lahko tudi pri ustvarjanju ponarejenih likov [40], kot je to v primeru filmu Rogue One: Zgodba Vojne zvezd (angl. Rogue One: A Star Wars Story), v katerem je igralec, ki je igral vlogo Grand Moffa Tarkina, umrl, vendar je bil njegov lik naknadno digitalno poustvarjen (slika 3.1). V istem filmu so uporabili tehnologijo globokih ponaredkov še za manipulacijo s starostjo igralcev, kot je primer igralke Carrie Fisher, vidno na sliki 3.2. Čeprav ti podvigi niso

bili popolni in je del občinstva opazil, da je na njih nekaj nenavadnega, je tehnologija od takrat zelo napredovala in je močno uporabljena v današnji kinematografski industriji.



Slika 3.1: Ponarejen (levo) in dejanski lik Grand Moffa Tarkina.

3.2 Jezikovne ovire

Druga podobna uporaba tehnologij globokih ponaredkov je bila v politični kampanji, ki je potekala v Indiji, kjer je vodja stranke Delhi Bharatiya Janata uporabil tehnologijo Deepfake za deljenje oglasa v angleščini [18]. Ker Janata ni govoril tega jezika, so dali igralcu prebrati sporočilo in uporabili globoko ponarejene tehnologije, da bi se zdelo, kot da vodja stranke dejansko govoriti v angleščini. S to potezo so uspeli približati volivce temu kandidatu z izločitvijo jezikovne ovire. Drug projekt, pri katerem je bila uporabljena podobna tehnologija globokih ponaredkov, je bila kampanja Malaria mora



Slika 3.2: Ponarejen (levo) in dejanski lik igralke Carrie Fisher.

umreti (angl. *Malaria must die*) [28], ki je bila sprožena z videoposnetki Davida Beckhama, nekdanjega angleškega profesionalnega nogometnika, ki je govoril devet jezikov. Medtem ko je v resnici govoril samo angleško, se v videou zdi, kot da dejansko govori ostalih osem jezikov, saj so se njegove ustnice premikale glede na besede, ki jih je govoril, čeprav je bil to glasovni posnetek. To je primer, da je mogoče tehnologije globokih ponaredkov uporabiti pri odpravljanju jezikovnih ovir. Posledično je lahko vsaka vsebina v katerem koli jeziku na voljo širšemu občinstvu, ne pa le govorcem izvirnega jezika.

3.3 Izobrazba

Tehnologije globokih ponaredkov bi lahko izboljšale izobraževalni proces z oživitvijo zgodovinskih osebnosti in odprle pot k interaktivnosti pri številnih predmetih. To idejo so že uresničili v muzeju in izobraževalnem centru holokavsta v Illinoisu (angl. Illinois Holocaust Museum) [25], kjer je organizacija naredila hologramske intervjuje z žrtvami holokavsta [27]. Tehnologija namreč deluje tako, da hologramska figura najprej pove vnaprej posneto zgodbo

lika, ki ga predstavlja, nato pa od obiskovalcev zahteva, da z vprašanji izvejo več. Nato algoritem z uporabo razpozname govora in tehnologije jezikovnih modelov, ki so naučeni s pomočjo nevronskih mrež in globokega učenja, pobere ključne besede iz vprašanja in nanj odgovori, kot da bi hologram dejansko govoril [3]. Drug zanimiv primer je razstava Dalí Lives v muzeju Dalí [20] v St. Petersburgu na Floridi. Sam Dalí je rekel: ”Na splošno verjamem v smrt, v Dalíjevo smrt pa nikakor ne.” Muzej je sprejel te besede kot izziv in ustvaril figuro Dalija v naravni velikosti, ki komunicira z obiskovalci tako, da jim pripoveduje zgodbe o njegovem življenju in na koncu z njimi posname tudi nekaj selfijev. To je bilo doseženo z lepljenjem Dalijevega obraza na moškega podobne konstitucije, ki je imitiral nekatere gibe. Glede barve glasu so najeli igralca, ki je znal posnemati Dalíjev naglas. Na koncu so pa video uskladili s pomočjo zvočnih posnetkov. Uporaba tehnologije globokih ponaredkov bi lahko odprla pot novemu izobraževalnemu pristopu, ki bi zbudil večje zanimanje za predmete pri učencih in jim dal možnost za bolj oseben stik z vsebino, ki se jo učijo.

3.4 Nevarnosti tehnologij globokih ponaredkov

Medtem ko smo videli, da lahko obstaja več uporab tehnologij globokih ponaredkov, ki so koristne za našo družbo, ta najsodobnejša tehnologija prinaša tudi številne nevarnosti. Zloraba tehnologije globokih ponaredkov ima lahko za posledico široko paletu primerov – od spletnega nasilja in agentov, do popolnih nacionalnih škandalov ter mednarodne krize in provokacij.

3.4.1 Izsiljevanje

Tehnologija globokih ponaredkov ponuja veliko različnih orodij, ki se lahko uporablja v namene izsiljevanja. Ko posameznik postane tarča izsiljevalcev, se lahko obraz žrtve nalepi na katero koli osebo na katerem koli videopo-

snetku, kjer počnejo vprašljive stvari, kar lahko vodi do obremenilnih posnetkov ljudi, ki dejansko niso storili ničesar podobnega. Tehnologija globokih ponaredkov je dragoceno orožje v rokah izsiljevalcev, kar je hkrati postala ena izmed največjih nevarnosti te tehnologije. Zanimiv pojav, ki se je zgodil kot posledica vse pogostejšega izsiljevanja, pa je takoimenovana "inflacija izsiljevanja" [35]. To je nov fenomen, kjer izsiljevani ljudje trdijo, da na videu niso oni in da je ta video rezultat globokega ponarejanja. Ta trditev se je začela uporabljati tudi, ko je posnetek izsiljevanja osebe avtentičen, zaradi česar je začela padati vrednost izsiljevanja (angl. blackmail inflation).

3.4.2 Vdori v medicini

Tehnologije globokih ponaredkov širijo svoj vpliv v medicinskom svetu, kot kaže raziskava CT-GAN: zlonamerno poseganje v 3D medicinske posnetke z uporabo globokega učenja [24]. V njem so avtorji izvedli uspešen napad na bolnišnico, kjer so spremenili 3D CT posnetke različnih pacientov z injiciranjem ali odstranitvijo pljučnega raka.

3.4.3 Pornografija in spolna zloraba

Po poročilu, ki ga je objavil Deeptrace, start-up podjetje za kibernetско varnost, je 96 odstotkov vseh globokih ponaredkov na spletu pornografskih. Večinoma subjekti v ogrožajočih videoposnetkih niso vedeli in niso dali svojega soglasja, da bi bili predstavljeni v takih videoposnetkih. Najpogosteje so cilj teh napadov znane ljudje, običajno ženske, npr. primer Scarlett Johansson [15], ki se je odločila, da posnetkov ne bo poskušala odstraniti, saj je to praktično nemogoče storiti, ko enkrat pridejo na internet. Drug primer slabe uporabe tehnologij globokih ponaredkov je aplikacija DeepNude [37], ki je bila odstranjena iz trgovine Google Play Store zaradi svoje kontroverzne narave. Ta aplikacije je izbrano sliko neke osebe spremenila tako, da je prikazala ponarejeno sliko, na kateri je bila ta ista oseba brez oblačil.

3.4.4 Računalniški agenti

S tehnologijo globokih ponaredkov se lahko ustvarja tudi ”nove” ljudi. S pomočjo globokih ponaredkov je mogoče ustvariti neobstoječ obraz osebe, ki se kasneje lahko uporabi za ustvarjanje spletnih profilov, pa tudi medijsko osebnost, ki jo je mogoče zlorabiti. Npr. v primeru Oliverja Taylorja [31], študenta iz Velike Britanije, ki je tudi samostojni novinar. V časopisih je objavil več kolumn o antisemitizmu in judovstvu. Tudi britanskega akademika Mazena Masrija in njegovo ženo Ryvko Barnard, ki je bila borka za pravice Palestincev, je obtožil, da sta znana simpatizerja teroristov. Ko so oblasti skušale od njega izvedeti kaj več, se je izkazalo, da ne obstaja. Njegova univerza ni imela podatkov o tem, da bi tam študiral, njegova telefonska številka je bila prekinjena, ni se oglasil na elektronsko pošto, ni imel internetnega odpisa in noben od časopisov, za katere je pisal, ga ni mogel identificirati. Slika, ki jo je uporabil, je bila deepfake in njegova identiteta še vedno ostaja neznana.

3.4.5 Politika

Videoposnetke, narejene s pomočjo globokih ponaredkov, je mogoče uporabiti tudi v politične namene, s ciljem, da bi dali izjavo, ali v propagandne namene. Takšen primer se je zgodil, ko je belgijsko okoljevarstveno gibanje Extinction Rebellion, ki uporablja kreativno nenasilno državljansko nepokorščino, da bi spodbudilo ukrepe proti klimatskim spremembam, posnelo globoko lažni videoposnetek belgijske premierke Sophie Wilmes [13]. V video ”premierka” govorji o krizi COVID-19 in jo povezuje s krčenjem gozdov. Video je hitro postal viralen in v prvem dnevu objave si ga je ogledalo več kot 100.000 ljudi. Številni ljudje, ki so ga videli, so si ga razlagali kot pristen video in ne kot globok ponaredek.

Poglavlje 4

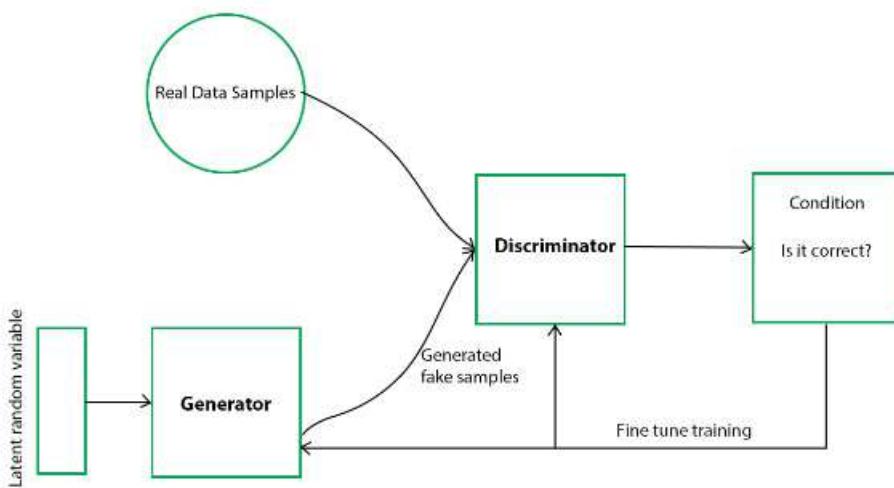
Pregled metod

Obstaja veliko število načinov za ustvarjanje globokih ponaredkov. V tem poglavju bomo preučili najpomembnejše metode, kot so Generativna nasprotniška mreža (angl. Generative Adversarial Network) – GAN, Variacijski samokodirnik (angl. Variational Auto-Encoders) – VAE, Variacijska generativna nasprotniška mreža samokodirnika (angl. Variational Auto-Encoders Generative Adversarial Network) – VAE-GAN, o katerih bomo govorili v tem poglavju. Večina jih temelji na samodejnih kodirnikih ali nevronskih mrežah.

4.1 Generativna nasprotniška mreža

Eden izmed najpogosteje uporabljenih pristopov je pristop GAN (angl. General Adversarial Network) [14]. V kontekstu tehnologij globokih ponaredkov se uporablja za izdelavo novih, neobstoječih obrazov. GAN je model globokega učenja, ki je sestavljen iz dveh nevronskih mrež, ki igrata igro z ničelno vsoto. Dve nevronski mreži tekmujeta ena proti drugi in izguba ene mreže je dobiček druge mreže in obratno. Imenujeta se generativna mreža (angl. Generator) in diskriminatorna mreža (angl. Discriminator). Generativna mreža generira nove primere (angl. Generated fake samples), medtem ko diskriminatorna mreža odloča, ali je vsak nov primer generiran ali resničen. Cilj generativne mreže ni ustvariti več resničnih primerov, temveč preslepiti

diskriminatorno mrežo. Funkcija, ki se uporablja v pristopu GAN, je funkcija Minimax, v kateri generator (generativna mreža) poskuša minimizirati ciljno funkcijo tako, da se na njej lahko izvaja gradientni spust, medtem ko diskriminatator (diskriminatorna mreža) poskuša maksimizirati ciljno funkcijo, tako da se na njem lahko izvaja strm vzpon. Celoten proces je razviden na sliki 4.1. GAN se uči s pomočjo izmeničnega vzpenjanja in spuščanja z gradientom, pri čemer daje prednost povečanju verjetnosti, da se diskriminatator moti, z izvajanjem vzpona z gradientom na generatorju na podlagi te ciljne funkcije. Pomanjkljivost GAN-a je, da je nestabilen za učenje, saj se obe omrežji učita iz enega samega povratnega širjenja, zaradi česar je zelo pomembna izbira ciljev.

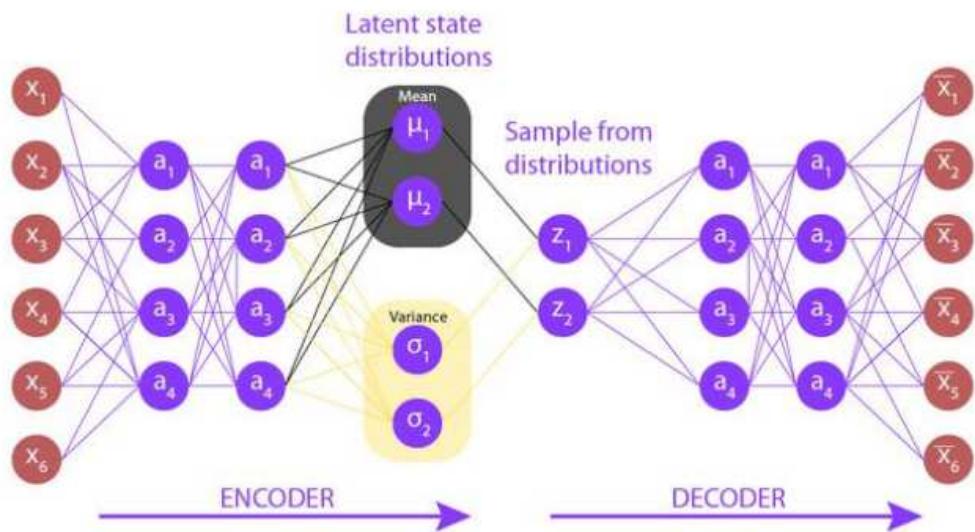


Slika 4.1: Generativna nasprotniška mreža (angl. Generative Adversarial Network) – GAN – Vizualna predstavitev [30].

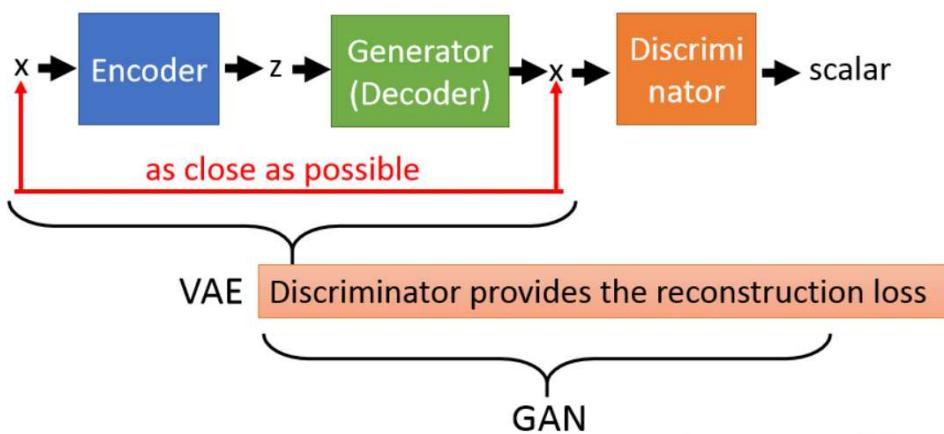
4.2 Variacijski samokodirnik

Druga pogosto uporabljena metoda za generiranje globokih ponaredkov je Variacijski samokodirnik (angl. Variational Autoencoder) – VAE [19]. Va-

riacijski samokodirnik je vrsta umetne nevronske mreže, ki je sestavljena iz kodirnika (angl. encoder) in dekodirnika (angl. decoder), oba pa se učita skupaj, da se zmanjša napaka pri rekonstrukciji med parametričnim in pravim aposteriorjem. Vizualizacija procesa delovanja VAE je prikazana na sliki 4.2. Njegovi vhodni podatki so ogroženi zaradi parametrizirane porazdelitve, kot je Bayesova porazdelitev. Slaba stran arhitekture VAE je zamegljenost izhodov, ki jih ustvari zaradi kombinacije obnovitve distribucije podatkov in funkcij izgube VAE. Nekatere arhitekture ta pristop pogosto uporabljajo v kombinaciji s pristopom GAN, kar nam daje Variacijske generativne nasprotniške mreže samokodirnika (angl. Variational Auto-Encoders Generative Adversarial Network) – VAE-GAN. Detaljni prikaz kombinacije VAE in GAN se lahko vidi na sliki 4.3. V VAE-GAN, dekodirnik (angl. Decoder) je v ulogi generatorja (angl. Generator), ter se trudi da je izvorna slika (x) podobna generirani (\tilde{x}).



Slika 4.2: Variacijski samokodirnik (angl. Variational Autoencoder) – VAE – Vizualna predstavitev [29].



Slika 4.3: Variacijske generativne nasprotniške mreže samokodirnika (angl. Variational Auto-Encoders Generative Adversarial Network) – VAE-GAN – Vizualna predstavitev [2].

Poglavlje 5

Pregled pristopov

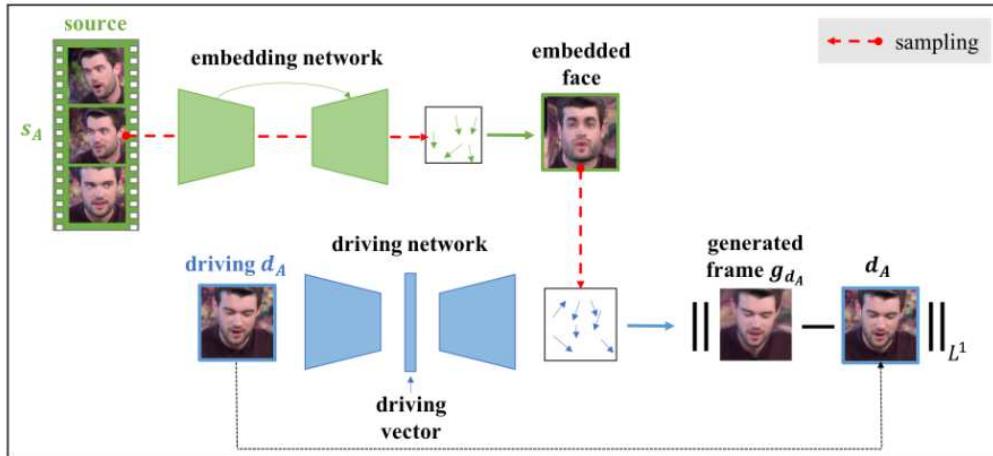
Metode, omenjene v poglavju 4, se uporabljajo z različnimi pristopi, kot so Nekajkratno učenje (angl. Few Shots Learning) – FSL, X2face, Model gibanja prvega reda (angl. First Order Motion Model) – FOMM in SimSwap, ki so razloženi v tem poglavju.

5.1 Nekajkratno učenje

Nekajkratno učenje (angl. Few Shots Learning) – FSL [41] je tehnika strojnega učenja, pri kateri se model usposobi za učenje iz majhnega števila primerov. Uporaben je v primerih, ko je za določeno nalogu na voljo omejena količina podatkov za učenje. Nekajkratni učni algoritmi lahko posplošijo iz majhnega števila primerov na nove, nevidene primere z uporabo predhodnega znanja, pridobljenega z drugimi nalogami. To modelu omogoča hitro in učinkovito učenje novih nalog, brez potrebe po veliki količini označenih podatkov. Nekatere pogoste tehnike, ki se uporabljajo pri nekajkratnem učenju, vključujejo transferno učenje, metaučenje in učenje na podlagi metrike.

5.2 Omrežje X2Face

Omrežje X2Face [39] je mreža, razvita za nadzor generiranja obrazov s pomočjo zvoka, slik in kod poze. Je model nevronske mreže, katerega cilj je dodati izvorni okvir pogonskemu okvirju. Izvorni okvir je ena ali več slik, ki zagotavljajo podrobnosti obraza oziroma, z drugimi besedami, obraz, ki ga želimo prikazati, medtem ko pogonski okvir prihaja iz pogonskega videoposnetka, ki vsebuje gibanje in pozno, za katero želimo, da je obraz iz izvirne slike. Uči se z velikim naborom video podatkov in je samonadzorovan. Metoda je sestavljena iz dveh podomrežij, ki za vhod sprejemata pare izvornih in pogonskih okvirjev. Dve podomrežji se imenujeta vgrajeno omrežje (angl. Embedding Network) in pogonsko omrežje (angl. Driving Network). Omrežje za vdelavo vzame izvorni okvir kot vhod in se nauči predstavitev vdelanega obrazu, zato se osredotoča na frontalizacijo obrazu. Pogonsko omrežje na drugi strani vzame pogonski okvir za vnos in se nauči preslikati iz vdelane predstavitev obrazu v ustvarjeni okvir s pomočjo pogonskega vektorja. Shema delovanja omrežja X2Face je razvidna na sliki 5.1.



Slika 5.1: Omrežje X2Face – Vizualna predstavitev [39].

5.3 Model gibanja prvega reda

Model gibanja prvega reda (Angl. First Order Motion Model) – FOMM [32] je pristop, ki združuje videz iz izvorne slike, kot so poteze obraza in pričeska, ter vzorce gibanja iz pogonskega videa. Učenje je doseženo s pomočjo velike zbirke videotrenutkov iste kategorije, zato bo model, ko bo usposobljen, deloval za vse slike iste kategorije. Metodo, ki jo uporablja ta pristop, je mogoče opisati s kodiranjem gibanja s pomočjo premikov ključnih točk, specifičnih za gibanje, in lokalnih afinih transformacij. Model deluje s pari izvorne slike in vsakega od okvirjev iz pogonskega videa kot vhoda. Nato se izvede animacija slike izvornega objekta. Ogrodje je sestavljeno iz dveh modulov – modula za oceno gibanja in modula za generiranje slike – in uporablja povratni optični tok. Najprej modul za ocenjevanje gibanja napove gosto polje gibanja iz okvira, nato pa modul za generiranje slike upodobi sliko izvornega predmeta, ki se premika s pomočjo pogonskega videa.

5.4 Metoda SimSwap

SimSwap [8] je pristop, ki je uporabljen v tej diplomski nalogi. To je ogrodje za zamenjavo obrazov v video sekvenkah, ki uporablja kombinacijo globokega učenja in tradicionalnih tehnik obdelave slik. Avtorji predlagajo cevovod, ki je sestavljen iz več ključnih komponent.

5.4.1 Pridobivanje obraznih potez

Globoka nevronska mreža se uporablja za pridobivanje obraznih potez, kot so obrazne mejnike in teksture iz izvornega in ciljnega obraza.

5.4.2 Poravnava obraza

Iзвlečene poteze obraza se uporabijo za poravnavo izvornega in ciljnega obraza, tako da se njihove obrazne točke in teksture ujemajo.

5.4.3 Deformacija obraza

Ta modul se uporablja za preoblikovanje izvornega obraza, da se ujema s ciljnimi obrazom. Ta korak vključuje ukrivljanje izvornega obraza, da se ujema z obliko in teksturo ciljnega obraza, hkrati pa upošteva gibanje obraza v videu.

5.4.4 Zamenjava obraza

Zadnji korak vključuje združitev preoblikovanega izvornega obraza s ciljnim obrazom za ustvarjenje realistične zamenjave obraza.

Avtorji trdijo, da lahko predlagani cevovod doseže visoko vizualno kakovost in naravnost v zamenjavah obrazov, ki jih ustvarja, hkrati pa je tudi računsko učinkovit. Omenjajo tudi, da uporabljajo semantično segmentacijsko mrežo za pridobivanje območja obraza, ki se nato posreduje v cevovod.

Če povzamemo, to ogrodje uporablja globoko nevronsko mrežo za pridobivanje obraznih potez, poravna in ukrivi izvorni obraz, da se ujema s ciljnim obrazom, nato pa oba obraza združi, da ustvari realistično zamenjavo obrazov v video zaporedju, kar je tudi računsko učinkovito.

Uporabo metode SimSwap v tej nalogi si bomo podrobnejše ogledali v poglavju 8.

Poglavlje 6

Pregled programske opreme

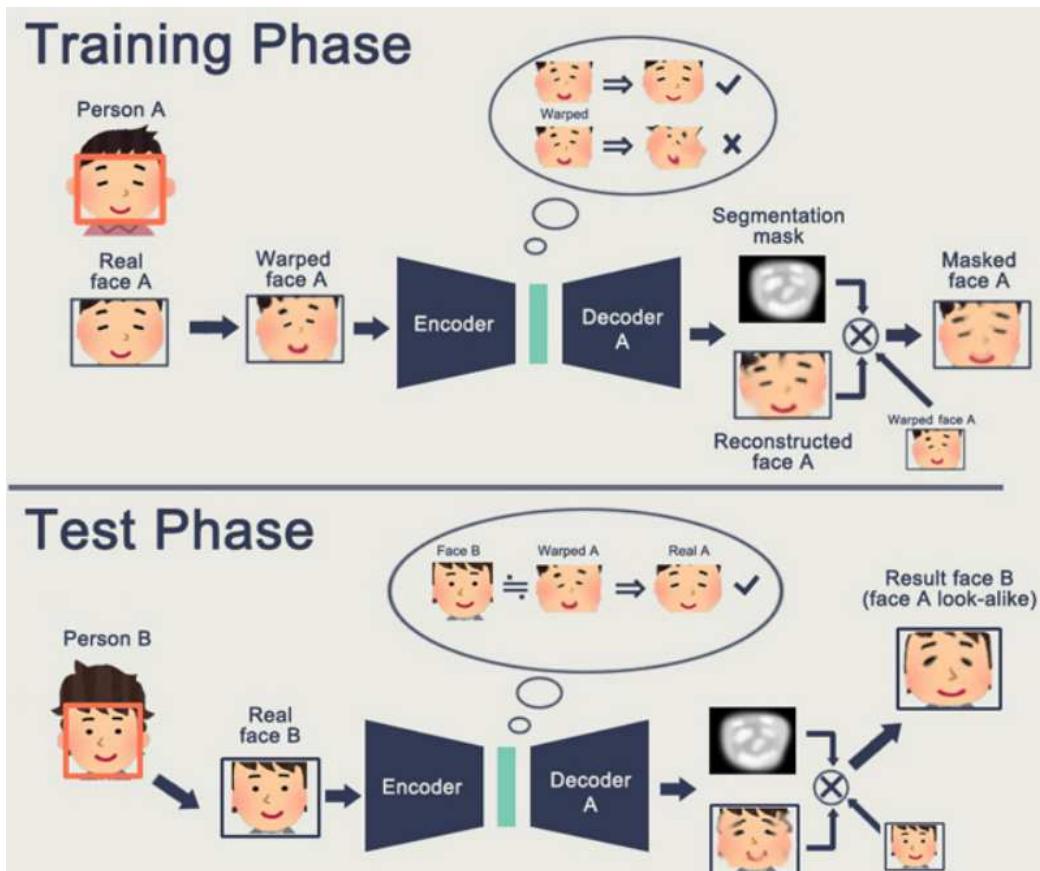
Na tržišču obstaja veliko programske opreme za ustvarjanje globokih ponaredkov. Sprva se je tehnologija razširila na platformi GitHub in je bila omejena za operacijski sistem Linux. Danes pa dobimo že samostojne programe z grafičnimi vmesniki, ki tečejo na različnih platformah. V nadaljevanju bomo predstavili najbolj razširjeno programsko opremo za ustvarjanje globokih ponaredkov: DeepFakes Faceswap, DeepFaceLive in Reface.

6.1 DeepFakes Faceswap

DeepFakes Faceswap [11] uporablja algoritme strojnega učenja za zamenjavo obraza ene osebe v videu ali sliki z obrazom druge osebe. Program uporablja tehnikе računalniškega vida in globokega učenja za analizo točk obeh obrazov in ustvarjanje nemotenega spajanja. Pristop za to je uporaba kodirnika in dveh identitetskih dekodirnikov za ustvarjanje novega obrazu, ki je kombinacija dveh vhodnih obrazov. Med učenjem (angl. Training Phase) se kodirnik in dekodirniki uporabljajo za povrnitev ukrivljenih slik (angl. warped face) v prvotno stanje. Enak postopek se uporablja za ciljne slike. Med testiranjem (angl. Test Phase) se kodirnik in dekodirniki uporabljajo za izdelavo slike, ki združuje identiteto izvorne slike z značilnosti ciljne slike s pomočjo ukrivljenih slik. Delovanje programa DeepFakes Faceswap se lahko vidi na

slike 6.1.

Program se običajno opira na tehnike zaznavanja obraznih mejnikov in poravnave obraza, da natančno prepozna in ujema ustrezne obrazne poteze, preden jih zamenja. Program se uči na naboru slik obrazov uporabnika, ki jih lahko pridobi tudi od videoposnetka, in ciljnega videa, da se izboljša njegovo delovanje in natančnost. Za dobre rezultate traja učenje več kot 24 ur. Končni rezultat je lahko v obliki videa ali slike.



Slika 6.1: Način delovanja programa DeepFakes Faceswap [22].

6.2 DeepFaceLive

DeepFaceLive [16] je sistem, ki uporablja tehnike globokega učenja za omogočanje menjave obrazov v realnem času v video tokovih v živo. DeepFaceLive je novejša različica programa DeepFaceLab, od katerega se razlikuje predvsem v tem, da DeepFaceLive dela v realnem času. Ta sistem je zgrajen po vzoru obstoječega ogrodja globokega učenja, kot je Tensorflow ali Pytorch, in uporablja vnaprej usposobljene modele, kot sta OpenFace ali FaceNet za prepoznavanje in poravnava obraza. Ko so obrazi zaznani in poravnani, sistem uporablja nevronsko mrežo za izvedbo zamenjave obrazov v realnem času.

DeepFaceLive uporablja Konvolucijsko nevronsko mrežo (angl. Convolutional Neural Network) – CNN za analizo točk obeh obrazov in ustvarjanje novega obraza, ki je kombinacija obeh vhodnih obrazov. Sistem je optimiziran za delovanje v realnem času in lahko deluje na različnih platformah.

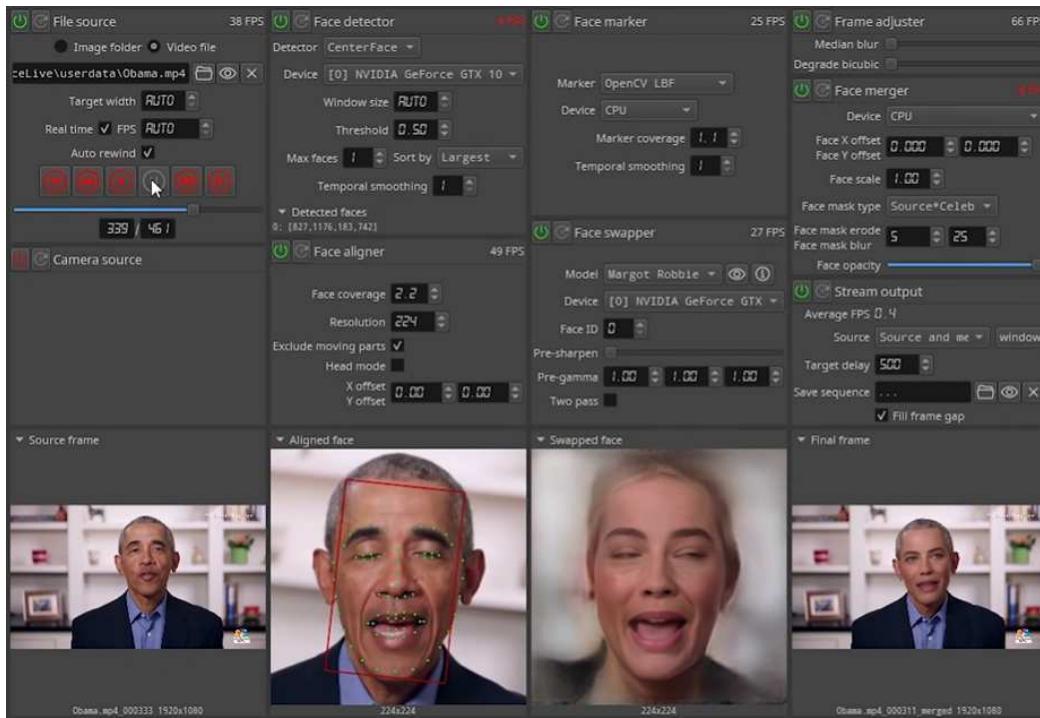
DeepFaceLive podpira učenje izvornega obraza, ki se potem lahko uporabi pri zamenjavo obraza nekega lika v realnem času. Vedež programa je viden na sliki 6.2.

Sistem lahko vključuje tudi dodatne funkcije, kot je možnost sledenja več obrazom hkrati, podpora za več vhodov kamere ter možnost snemanja in shranjevanja izhodnega videa. Pomanjkljivost je, da potrebuje učenje na izvornem obrazu in da celotno učenje traja predolgo, kar pomeni, da ni uporaben za hitro zamenjavo obraza naključne izvorne slike osebe.

6.3 Reface

Reface[26] je mobilna aplikacija, ki uporablja algoritme strojnega učenja za zamenjavo obraza osebe v videu ali sliki z obrazom uporabnika. Aplikacija uporablja tehnologijo za prepoznavanje obraza in sledenje, da natančno identificira in poravna uporabnikov obraz z obrazom na videu ali sliki. Ta postopek poravnave je ključen za doseganje realističnega in brezhibnega spašanja obeh obrazov.

Aplikacija se običajno opira na globoke nevronске mreže, kot so konvolu-



Slika 6.2: Vedež program DeepFaceLive.

cijanske nevronske mreže (CNN) ali generativne nasprotniške mreže (GAN), za analizo značilnosti obeh obrazov in ustvarjanje novega obraza, ki je kombinacija uporabnikovega obraza in obraza v videu ali sliki. Te nevronske mreže so vnaprej usposobljene na velikem naboru podatkov slik obrazov iz videa in natančno nastavljene za izboljšanje učinkovitosti in natančnosti postopka zamenjave obrazov.

Končni rezultat aplikacije je video ali slika z obrazom uporabnika, zamenjamim z obrazom na videu ali sliki. Pomanjkljivost te aplikacije je, da ni odprtnokodna ter se ne da učiti poljubnega videa. Kot vhod aplikacija prejme sliko uporabnika in generira rezultate v roku 1 minute za video s trajanjem do 1 minute. Ne obstaja opcija samostojnega učenja, uporabnik pa lahko izbira zamenjavo izbranega obraza le za obraze iz že obdelanega video posnetka.

Poglavlje 7

Metoda SimSwap

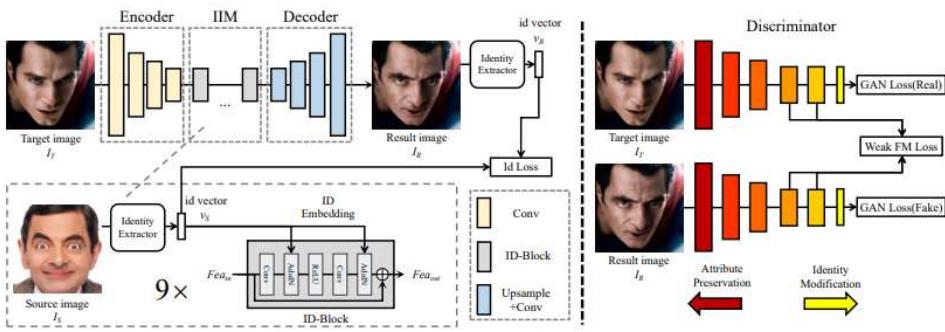
Ogrodje SimSwap [8] je orodje, ki se uporablja za prenos identitete izvornega obraza v identiteto ciljnega obraza. V tem procesu okvir ne spremeni značilnosti ciljnega obraza. Metoda SimSwap je specifična v tem, da ne gre za arhitekturo, ki je specifična samo za eno identiteto zamenjave, temveč se njene funkcionalnosti razširijo na zamenjavo poljubne identitete. Način delovanja ogrodja SimSwap se lahko razdeli v 3 ključne dele: arhitektura, na kateri temelji metoda, zamenjava poljubne identitete ter funkcija izgube.

7.1 Arhitektura, na kateri temelji metoda

Ogrodje SimSwap temelji na metodi zamenjave obrazov, ki se uporablja v programski opremi DeepFakes Faceswap, opisane v poglavju 6. Glavna omejitev metode, uporabljene v DeepFakes Faceswap je, da jo je mogoče uporabiti samo za eno identitet. To je zato, ker metoda zahteva, da so informacije o identiteti izvornega obrazu integrirane v uteži enega od dekodirnikov metode. SimSwap presega to omejitev z razširitvijo funkcionalnosti metode, da lahko zamenja poljubno identitet.

7.2 Zamenjava poljubne identitete

Da bi se izognili omejitvam, opisanim v prejšnjem razdelku, poskuša metoda SimSwap ločiti identitet iz dekodirnika. Ta korak omogoča uporabo metode SimSwap s poljubnimi identitetami. To je doseženo z uvedbo dodatnega modula za vstavljanje identitete (angl. ID Injection Module , IIM) med kodirnikom in dekodirnikom. Celotna arhitektura se lahko vidi na sliki 7.1.



Slika 7.1: Metoda SimSwap [8].

Po pridobivanju značilnosti ciljne slike (I_T) prek kodirnika je naš cilj zamenjati ciljno podobo z obrazom izvirne slike (I_S). Za izpolnjenje te naloge je treba podatke o identiteti ciljne slike nadomestiti z informacijami identitete izvornega obraza, pri tem pa je tudi treba ohraniti podatke o drugih značilnosti ciljnega obraza nespremenjene. Vendar pa je razlikovanje med seboj zelo povezanih informacij o identiteti in ostalih značilnosti ciljne slike velik izliv. Zato se spremembu izvede na vseh značilnostih ciljne slike ($FeaT$), pri čemer se metoda zanaša na učno izgubo (angl. training loss). To pomaga, da mreža med učenjem loči, katere značilnosti spremeniti in katere obdržati.

Modul za vstavljanje identitete (IIM), ki je bil dodan, je orodje, ki spremeni podatke o identiteti v značilnostih ciljnih slik v podatke o identiteti izvornega obraza. Sestavljen je iz dveh delov: dela za pridobivanje identitete in dela za njeno vdelavo.

V delu pridobivanja identitete se obdela vhodna izvorna slika (I_S), ki

vsebuje podatke o identiteti in značilnostih izvornega obraza. Ker pa modul za vstavljanje ID-ja potrebuje samo podatke o identiteti, se za pridobivanje vektorja identitete iz izvorne slike (I_S) uporabi mrežo za prepoznavo obrazov ArcFace [9].

V delu za vdelavo se uporablja ID-blok (angl. ID-Block) za vnos pridobljenih informacij o identiteti v značilke. ID-blok je spremenjena različica residualnega bloka (angl. residual block), ki se pogosto uporablja v arhitekturah globokega učenja. Namesto paketne normalizacije (angl. batch normalization) se uporablja prilagodljiva normalizacija primerkov (angl. Adaptive Instance Normalization, AdaIN). AdaIN je normalizacijska tehnika, ki normalizira povprečje in standardni odklon funkcij na podlagi informacij o identiteti. Uporaba AdaIN v SimSwap se lahko definira z enačbo 7.2.

$$\text{AdaIN}(Fea, v_S) = \sigma_S \frac{Fea - \mu(Fea)}{\sigma(Fea)} + \mu_S$$

Slika 7.2: Enačba AdaIN.

V enačbi sta $\mu(\text{Fea})$ in $\sigma(\text{Fea})$ povprečje in standardni odklon vhodnih značilnosti, izračunani po posameznih kanalih (angl. channel-wise), ter sta generirana iz vektorja identitete izvorne slike (V_S) z uporabo polnopovezane mreže. Za zagotovitev zadostne informacije glede identitete se skupaj uporabi devet takšnih ID-blokov.

Po uporabi modula za vstavljanje identitet, ki spremeni značilnosti ciljnega obraza, da se ujema z identiteto izvornega obraza, se spremenjene značilnosti prenesejo skozi dekodirnik, da se ustvari končna slika (I_R). Ker so v fazu učenja vključene izvorne slike različnih identitet, uteži dekodirnika niso povezane samo z določeno identiteto. Dekodirnik je usposobljen za obnovitev slike iz spremenjenih značilnosti, ne da bi spremenil podatke o identiteti. To omogoča, da se arhitektura uporabi za poljubne identitete, namesto da bi bila omejena na določene identitete.

Med učenjem se izguba identitete (angl. identity loss) uporablja za zmanjšanje

razdalje med pridobljenim vektorjem identitete iz ustvarjene slike (V_R) in vektorjem identitete izvornega obraza (V_S). Vendar lahko zmanjševanje izgube identitete privede do prekomernega prilagajanja mreže (angl. overfitting), kjer mreža generira samo frontalne obraze z identitetom izvorne slike obraza, medtem ko izgubi vse značilnosti ciljnega obraza. Da bi se izognili tej težavi, se uporablja nasprotniško učenje (angl. adversarial training) z uporabo diskriminatorja za razlikovanje med ustvarjenimi slikami z očitnimi napakami. Nasprotniška izguba (angl. adversarial loss) se uporablja za izboljšanje kakovosti ustvarjenih slik. Uporablja se različica diskriminatorja PatchGAN, ki je vrsta nevronske mreže, ki se osredotoča na lokalne regije slike in ne na celotno sliko [17].

7.3 Funkcija izgube

Funkcija izgube, ki se uporablja v metodi SimSwap, je sestavljena iz petih komponent: izguba identitete (angl. identity loss), označena z L_{ID} , izguba pri rekonstrukciji (angl. reconstruction loss), označena z L_{Recon} , nasprotniška izguba (angl. adversarial loss), označena z L_{Adv} , gradientna kazen (angl. gradient penalty), označena z L_{GP} , in izguba ujemanja šibkih značilnosti (angl. weak feature matching loss), označena z L_{wFMsu} . Za nasprotniško izgubo se uporablja izguba Hinge (angl. hinge loss) [21]. Za boljšo učinkovitost pri velikih nefrontalnih položajih glave se uporablja večstopenjske diskriminatorenje (angl. multi-scale discriminator) [38]. Za preprečitev gradientne eksplozije se uporablja gradientno kazen [4]. Celotna funkcija izgube je opisana z enačbo 7.3

$$L = 10L_{Id} + 10L_{Recon} + L_{Adv} + 10^{-5}L_{GP} + 10L_{wFMsu}$$

Slika 7.3: Enačba za celotno izgubo.

7.3.1 Izguba identitete

Izguba identitete se uporablja za omejitev razdalje med vektorjem izvirne slike (V_S) in vektorjem končne slike (V_R). Za izračun razdalje se uporablja kosinusno podobnost (formula 7.4).

$$L_{Id} = 1 - \frac{v_R \cdot v_S}{\|v_R\|_2 \|v_S\|_2}$$

Slika 7.4: Enačba za izgubo identitete.

7.3.2 Izguba pri rekonstrukciji

Če sta izvorni in ciljni obraz iste identitete, mora biti ustvarjeni rezultat videti enako kot ciljni obraz. Če pa sta izvorni in ciljni obraz iz različnih identitet, je izguba pri rekonstrukciji nastavljena na 0. Enačbo za izgubo pri rekonstrukciji predstavimo z enačbo 7.5, kjer I_R predstavlja ustvarjeni rezultat, I_T pa ciljno sliko.

$$L_{Recon} = \|I_R - I_T\|_1$$

Slika 7.5: Enačba za izgubo pri rekonstrukciji.

7.3.3 Izguba ujemanja šibkih značilnosti

Izguba ujemanja šibkih značilnosti se uporablja za ohranjanje značilnosti ciljnega obraza. Pri zamenjavi obraza bi morale biti spremembe omejene samo na identiteto, značilnosti ciljnega obraza pa naj bi ostale enake. Ker SimSwap dela direktnе spremembe na vseh značilnostih ciljnega obraza, kar vsebuje tako identiteto kot značilnosti ciljnega obraza, predstavljena funkcija izgube ohrani določene značilnosti. Vendar to ne naredi za posamezno

značilnost posebej, ampak za vse hkrati s tem, da vzame samo manjši gornji del nivojev. Funkcija je zasnovana na originalni enačbi izgube ujemanja značilnosti ($L_{oFM}(D)$), prikazani z enačbo 7.6, kjer je $D^{(i)}$ nivo številke i od diskriminatoryja D , $N^{(i)}$ je število elementa na nivoju i , M je število nivojev v diskriminatoryju, I_R je ustvarjeni rezultat in I_{GT} je referenčna vrednost (angl. ground truth).

$$L_{oFM}(D) = \sum_{i=1}^M \frac{1}{N_i} \|D^{(i)}(I_R) - D^{(i)}(I_{GT})\|_1$$

Slika 7.6: Enačba za originalno izgubo ujemanja značilnost.

Enačba, ki opisuje izgubo ujemanja šibkih značilnosti ($L_{wFM}(D)$) pa je razvidna iz enačbe 7.7. Razlici med originalno enačbo izgube ujemanja značilnosti in enačbo za izgubo ujemanja šibkih značilnosti sta v tem, da enačba za izgubo ujemanja šibkih značilnosti uporablja le nivoje diskriminatoryja, ki so višji kot nivo m ter namesto referenčne vrednosti I_{GT} uporablja ciljno sliko I_T . Ker se uporablja večstopenjski diskriminator (angl. multi-

$$L_{wFM}(D) = \sum_{i=m}^M \frac{1}{N_i} \|D^{(i)}(I_R) - D^{(i)}(I_T)\|_1$$

Slika 7.7: Enačba za izgubo ujemanja šibkih značilnosti.

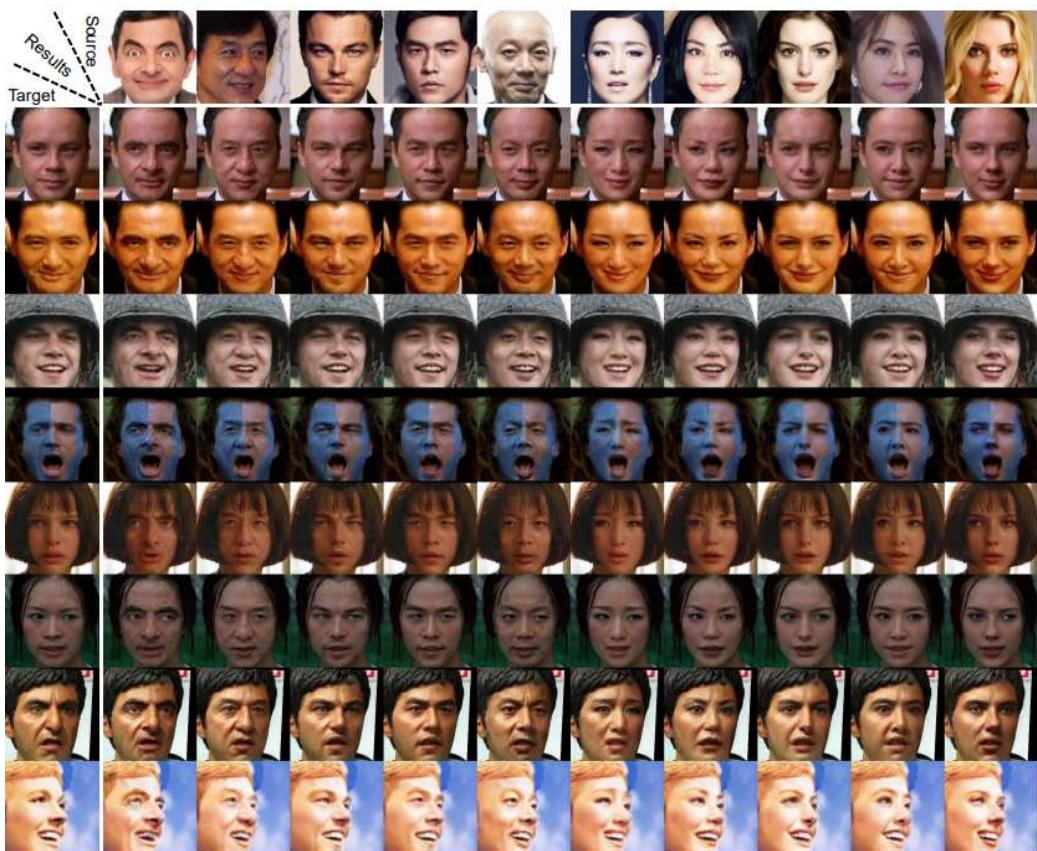
scale discriminator), je izguba ujemanja šibkih značilnosti izračunana za vse diskriminatoryje po enačbi 7.8.

$$L_{wFMSum} = \sum_{i=1}^2 L_{wFM}(D_i)$$

Slika 7.8: Enačba za izgubo ujemanja šibkih značilnosti.

7.4 Rezultati

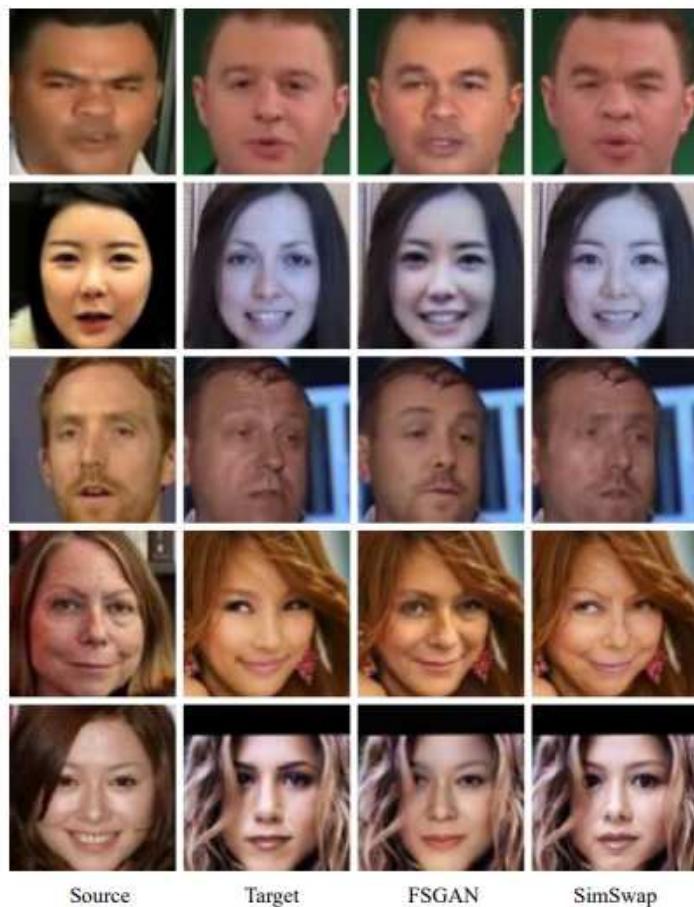
Metoda SimSwap je sposobna prenesti identiteto izvornega obraza v ciljni obraz, pri tem pa ohraniti značilnosti (kot so izraz, smer pogleda, drža in svetlobni pogoji) ciljnega obraza. SimSwap lahko obravnava različne identitete, vključno s težkimi ciljnimi pogoji, kot so pretiran izraz, črte na obrazu in velike rotacije obraza, kot je lahko vidno na sliki 7.9.



Slika 7.9: Matrika obrazov narejena z metodo SimSwap [8].

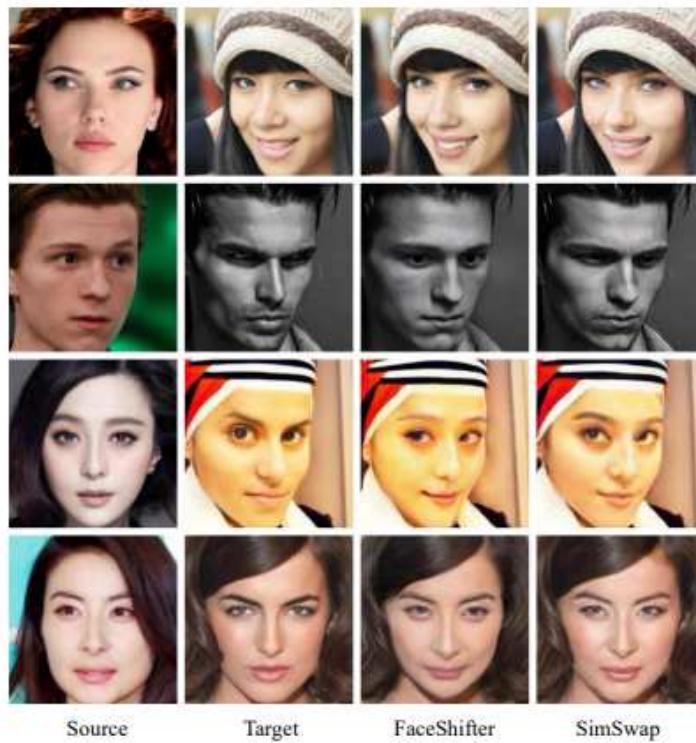
Da bi se ocenilo delovanje metode SimSwap, so jo avtorji primerjali s tremi drugimi metodami zamenjave obraza: FSGAN, DeepFakes Faceswap in FaceShifter. SimSwap prekaša DeepFakes in FSGAN pri ohranjanju značilnosti in dobrih rezultatih zamenjave identitete. Primeri zamenjave obrazov s FSGAN

in SimSwap so razvidni na sliki 7.10, kjer je narejena vizualna primerjava teh dveh metod. Prav tako je metoda SimSwap primerjana z metodo FaceShifter, ki ima močno sposobnost spremjanja identitete, vendar pogosto ne obdrži značilnosti, kot sta izraz in smer pogleda. SimSwap dosega boljše delovanje kot FaceShifter pri ohranjanju značilnosti, vključno z izrazom in svetlobnimi pogoji. Primerjava med metodama SimSwap in Faceshifter je vidna na sliki 7.11. Poleg te primerjave je na sliki 7.12 prikazana primerjava metod SimSwap, DeepFakes Faceswap in Faceshifter.



Slika 7.10: Primerjava metode SimSwap z metodo FSGAN [8].

Za namene primerjave rezultatov sta bile naučeni še dve mreži; SimSwap-



Slika 7.11: Primerjava metode SimSwap z metodo Faceshifter [8].

oFM in SimSwap-nFM. SimSwap-oFM uporablja izvirno formulacijo za ujemanje značilnosti, SimSwap-nFM pa ne uporablja ujemanja značilnosti. Poleg vizualne primerjave v članku [8] predstavijo tudi kvalitativne preizkuse za primerjavo s temo dvema spremenjenima metodama SimSwap ter z metodo DeepFakes Faceswap in Faceshifter. Primerjani kakovosti sta identiteta in poza. Za primerjavo so naključno izbrali 10 okvirjev iz različnih video-posnetkov z različnimi ljudmi iz baze FaceForensics++. Teh 10 okvirjev so potem razporedili v pare za zamenjavo obrazov, kjer je en izvorni, drugi pa ciljni obraz. Obrazi so bili zamenjani z uporabo metode SimSwap. Poleg tega so s pomočjo druge mreže za prepoznavanje obrazov pridobili identifikacijske vektorje za ciljne in generirane okvire. Za določitev parametrov identitet je uporabljen pristop, v katerem se za vsak zamenjan obraz išče obraz iz originalnih okvirjev, ki mu je najbolj podoben. Identiteta se je merila z na-



Slika 7.12: Primerjava z Deepfakes Faceswap in Faceshifter [8].

tančnostjo pogoja, če je ta obraz tisti, ki se je uporabljal kot izvorni obraz pri zamenjavi. Če je obraz, ki je najbolj podoben zamenjanemu obrazu, izvorni obraz, ki se je uporabil za zamenjavo, se natančnost poveča. To natančnost v tabeli 7.1 označuje parameter identiteta.

Parameter poza se pridobi s primerjanjem poze v originalnem in ustvarjenem okviru. Predstavljen je kot razdalja med pozno obrazu v originalnem in ustvarjenem okviru. Rezultati obeh parametrov za različne modele so vidni v tabeli 7.1.

Glede na primerjavo z ostalimi metodami je razvidno, da je metoda SimSwap hitra in učinkovita metoda zamenjave obraza, ki daje rezultate visoke podobnosti z izvornim obrazom, hkrati pa ohranja lastnosti ciljnega obraza. SimSwap prekaša druge metode zamenjave obraza pri ohranjanju značilnosti in je robustna pri zahtevnih ciljnih pogojih, vključno s pretiranim izrazom,

Metoda	Identiteta	Poza
DeepFakes Faceswap	77,65 %	4,59
Faceshifter	97,38 %	2,96
SimSwap-oFM	73,64 %	1,22
SimSwap	92,83 %	1,53
SimSwap-nFM	96,57 %	2,47

Tabela 7.1: Kvalitativni eksperimenti na bazi FaceForensics++ [8].

črtami na obrazu in velikimi zasuki obraza. Hkrati pa je tudi hitrejša in bolj praktična v primeru hitre zamenjave obraza, ker se model ne uči na izvornem obrazu.

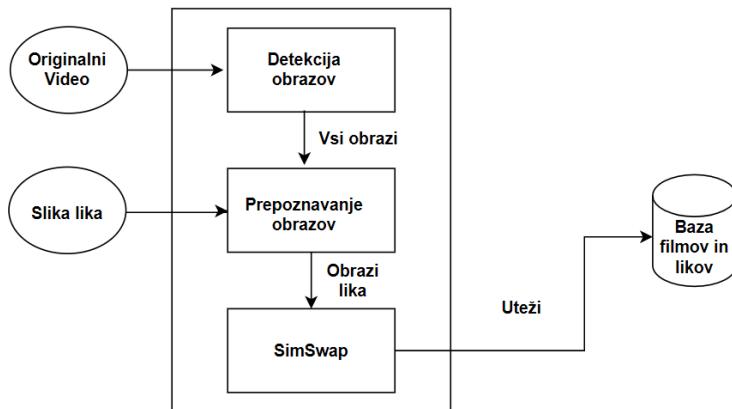
Poglavlje 8

Izdelava programa za zamenjavo obraza

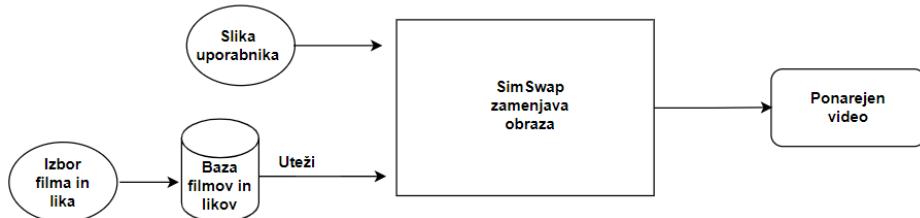
8.1 Ideja

Končni izdelek te diplomske naloge je program, napisan po načelih metode SimSwap [8]. Glavna ideja tega programa je, da lahko zamenja obraz svojega uporabnika in enega od glavnih junakov v nekaterih klasikah slovenske kinematografije. Da bi bil pripravljen za uporabo, ima program dva načina delovanja: način pripravljanja videa in način zamenjave obrazov. Del programa, namenjen za pripravljanje videa (slika 8.1), je sestavljen iz metode, ki za vhod vzame video posnetek slovenskega filma (ali kateri koli drug video) ter sliko lika iz filma, ki bo zamenjan, in izdela s pomočjo modela, uporabljenega v metodi SimSwap, uteži, ki se nato uporablja v delu programe, ki je namenjen zamenjavi obrazov. Del programa, namenjen zamenjavi obrazov (slika 8.2), je tisti del, ki je namenjen široki uporabi in je uporabniku prijaznejši. Program od uporabnika zahteva, da iz menija že pripravljenih videov izbere film in nato iz teh videov izbere lik. Nato je uporabnik pozvan, da zajame svojo sliko, ki bo uporabljena za zamenjavo obraza. Ko od uporabnika zbere vse te informacije, program nadaljuje z zamenjalnim delom, ki traja približno toliko dolgo, kot traja video. Na koncu program izdela in

shrani ponarejeni video, v katerem se obraz izbranega lika iz izbranega filma zamenja z obrazom uporabnika.



Slika 8.1: Shema delovanja dela programa, namenjenega za pripravo videa.



Slika 8.2: Shema delovanja dela programa, namenjenega zamenjavi obrazov.

8.2 Razlike med SimSwap in našim programom

Glavno težavo v povezavi s programom je predstavljal čas izvajanja. Če bi uporabili izvirno metodo SimSwap, bi bil čas izvajanja programa približno

10-krat daljši od dolžine izbranega videa. Ker smo želeli funkcionalen in hitter program, je bilo to nesprejemljivo. Rešitev te težave je bila sprememba metode SimSwap in njena razdelitev na dva dela; pripravljanje videa in zamenjava obraza. Del zamenjave obraza je uporabljal uteži, ki so bile shranjene v delu za pripravljanje videa. Program še vedno uporablja iste modele in principe kot SimSwap, vendar jih uporablja na časovno učinkovitejši način. S tem pristopom so uteži že dostopne, preden uporabnik zažene program, tako da so pripravljeni za uporabo v delu programa, ki je namenjen zamenjavi obrazov.

8.3 Pripravljanje videa

Prva faza programa je izbiranje in prepoznavanje obrazov v videu, na katerem se bodo izračunale uteži. Ta faza je sestavljena iz dveh delov: iskanje vseh obrazov v videu in zaznavanje, kateri od teh obrazov pripadajo isti osebi. V originalni kodi za SimSwap je obstajala možnost zamenjave enega, več ali vseh obrazov v videu, ki bi bile podane kot vhodne argumente v programu, vendar je za potrebe te diplomske naloge uporabljena le možnost zamenjave določenega obrazu. Ta del poteka tako, da vhodni video porazdeli na posamezne okvire ter v vsakem okvirju poišče vse obraze in jih označi. Potem preveri, ali je obraz lika, ki bo zamenjan, med prepoznanimi obrazi in če je, se ta okvir uporabi za pridobivanje uteži. Uteži, ki so nastale kot rezultat prvega dela programa, se shranijo za nadaljnjo uporabo pri zamenjavi obrazov.

8.3.1 Pridobitev obrazov

Za pridobitev vseh obrazov program uporablja ogrodje InsightFace [1], ogrodje za globoko učenje, ki se uporablja za nalogo prepoznavanja obrazov. InsightFace temelji na globokih nevronskeih mrežah, ki se učijo na velikih naborih slik obrazov. Omrežna arhitektura, uporabljena v InsightFace, je konvolucionska nevronska mreža (angl. Convolutional Neural Network – CNN). Omrežje

se uči z različnimi tehnikami, kot so povečevanje podatkov (angl. data augmentation), normalizacija serije (angl. batch normalization) in razpad teže (angl. weight decay), da se izboljša njegova natančnost in robustnost. Vsi obrazi, ki so pridobljeni v tej fazi, so potem v naslednjem koraku primerjani z obrazom enega izmed likov izbranega filma. Detektor obraza, ki je uporabljen v ogrodju InsightFace je RetinaFace [10].

8.3.2 Prepoznavava obrazov

Za primerjanje obrazov na videu program uporablja ArcFace [9]. ArcFace je metoda globokega učenja za prepoznavanje obrazov. Predstavljena je bila leta 2018 in je od takrat postala eden najbolj priljubljenih algoritmov za prepoznavanje obrazov. Metoda temelji na konvolucijskih nevronskeih mrežah (CNN) in uporablja funkcijo izgube, ki upošteva podobnost med značilnostmi in razlike med razredi za izboljšanje natančnosti prepoznavanja obrazov. Funkcija izgube v ArcFace je zasnovana za povečanje variance med razredi in zmanjšanje variance znotraj razreda, kar vodi k bolj diskriminativni predstavitev lastnosti. Funkcije, ki se jih nauči omrežje, se uporabljajo za izračun ocene podobnosti med vhodnim obrazom in vsakim obrazom v bazi podatkov. Najvišji rezultat pomeni najboljše ujemanje. Dokazano je, da se ArcFace dobro obnese pri nalogah prepoznavanja obrazov, vključno s preverjanjem obraza (določanje, ali dva obraza pripadata isti osebi) in identifikacijo obraza (ujemanje obraza z že obstoječo zbirko podatkov obrazov). Poleg tega je metoda robustna na variacije položaja ali osvetlitve. S pomočjo te metode se izloči vse obraze, ki pripadajo liku iz filma, katerega obraz bo zamenjan z obrazom uporabnika.

8.3.3 Organizacija podatkov

Za vsak video, na katerem se program uči, ustvari mapo, v kateri se nahajajo podmape za vsak lik iz videa, ki je bil že obdelan. V tej podmapi se nahajajo shranjeni podatki, ki se bodo uporabili, ko bo uporabnik izbral to video, in

pa lik za zamenjavo osebe.

8.4 Zamenjava obraza

Drugi del programa, ki je hkrati del namenjen širši javnosti, je menjava obrazov. Ta del lahko razdelimo na dva dela: prvega, s katerim pridobimo podatke o uporabniku, in drugega, v katerem poteka obdelava videa.

8.4.1 Zbiranje informacij

V tem delu programa je uporabnik pozvan, da izbere film in lik iz tega filma, s katerim bo zamenjal svoj obraz. Razpoložljivi videoposnetki in liki se preberejo iz mape, v kateri so shranjeni vsi predhodno usposobljeni modeli. Po izbiri videoposnetka in lika se uporabniku prikaže okno kamere, s katerim naj bi posnel sliko, ki bo uporabljena naprej v programu. Ko je slika pridobljena, program preide v drugo fazo – obdelavo videa.

8.4.2 Obdelava videa

Video obdelava najprej razdeli posnetek iz filma na zvočno komponento in okvirje. Nato razčleni okvirje. Če okvir ne vsebuje predhodno zapisanih informacij o liku, ki naj bi ga zamenjal uporabnik, potem je dodan takšen, kot je v predhodno pripravljenem nizu, namenjenim shranjevanju okvirjev. Če pa okvir vsebuje predhodno zapisane informacije o liku, se okvir obdela z nalaganjem uteži ter pridobivanjem podatkov o pridobljenem obrazu uporabnika. Ko je to končano, se obraz lika zamenja z obrazom uporabnika s pomočjo metode SimSwap in knjižnice OpenCV. Nov obraz se združi z okvirjem, okvir pa se doda nizu okvirjev. Ko so obdelani vsi okvirji, jih program združi v video format in doda zvočno komponento, ki jo je shranil v začetni fazi obdelave videa. Nastali videoposnetek se nato shrani in je pripravljen za ogled.

Lik	Okvirji	Pridobijeni obrazi	Prepoznani obrazi
Kekec	469	469 (100 %)	174 (37,10 %)
Pehta	459	459 (100 %)	21 (4,68 %)
Ivo	424	424 (100 %)	276 (65,09 %)
Mileva	505	505 (100 %)	353 (69,90 %)

Tabela 8.1: Analiza delovanja programa za zamenjavo obraza v primeru, ko je lik obrnjen na strani.

8.5 Primer delovanja

Primer delovanja programa za zamenjavo obraza si bomo pogledali na primeru filma ”Srečno, Kekec!”, slovenskem mladinskem filmu iz leta 1963, ki ga je režiral Jože Gale, in serije ”Bratovščina Sinjega galeba”, slovenske serije iz leta 1970, ki jo je režiral France Štiglic. V filmu ”Srečno, Kekec!” je ciljni video scena, v kateri sta dva lika – Kekec ter Pehta. V seriji ”Bratovščina Sinjega galeba” pa je ciljni video scena, v kateri so trije liki – Ivo, stric Just in Mileva. Program vpraša uporabnika za izbiro lika iz filma ter odpre okno s kamero, da bi ujel sliko uporabnika. Nato sledi izdelava globoko ponarejenega videa, ki je na koncu shranjen na računalnik uporabnika.

8.6 Analiza uspešnosti

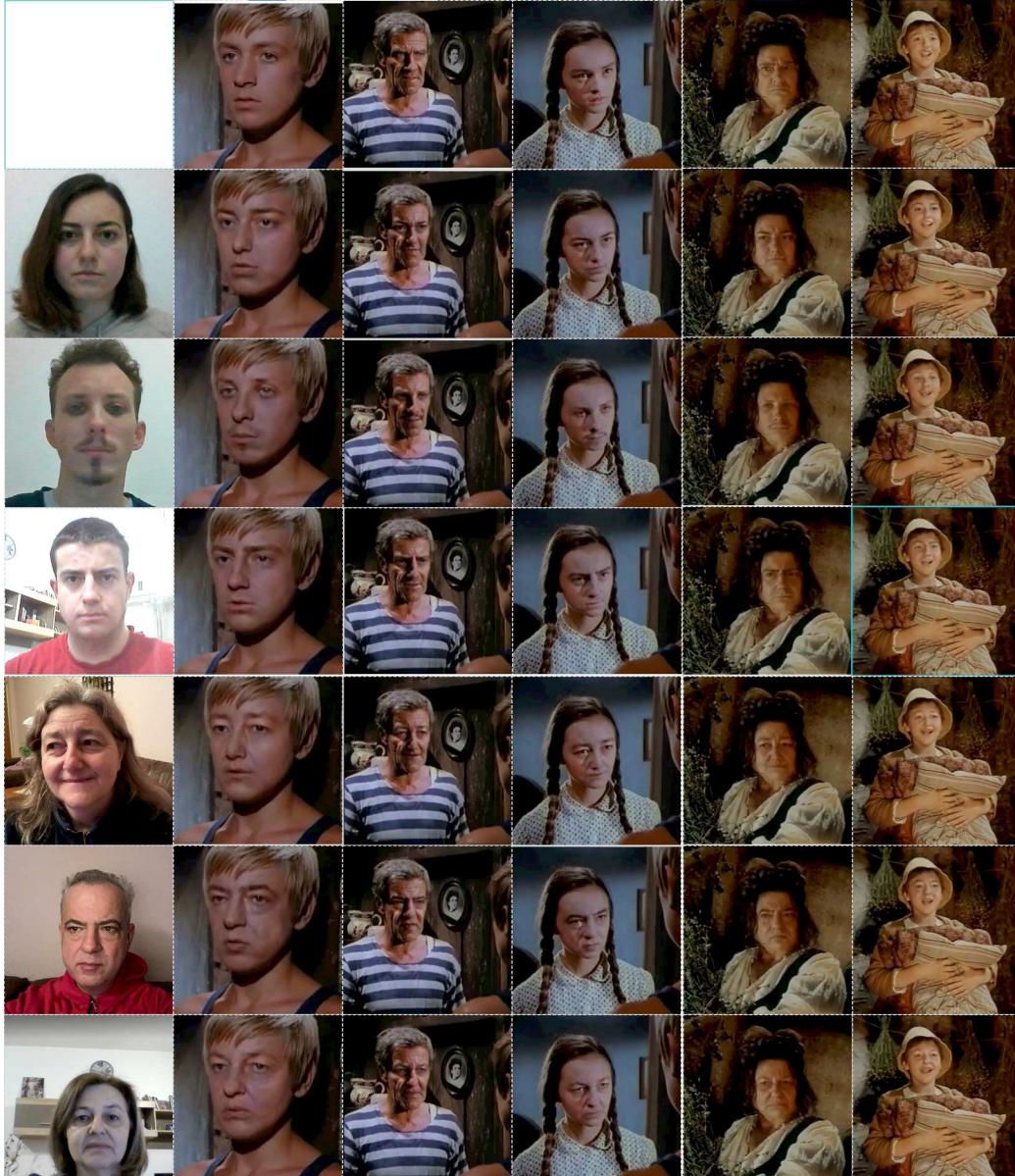
Programa za zamenjavo obraza daje odlične rezultate, če je lik obrnjen proti kamери. Primeri zamenjave obraza, ko je lik obrnjen proti kameri so vidni na sliki 8.3.

Težava je, ko je lik zajet iz strani in v tem primeru se zamenjava ne zgodi. Da bi ugotovili kje in kako pogosto se ta napaka zgodi, so bili vzeti in analizirani 15-sekundni vzorci videoposnetkov, ki vsebujejo obraz lika v vsakem okvirju. Rezultati analize so razvidni v tabeli 8.1.

Bili so prešteti okvirji (v tabeli 8.1: Okvirji), na katerih se dejansko pojavi

obraz lika. Ker so vzorci bili vzeti tako, da je lik na vsakem okvirju, je ta številka enaka številu okvirjev v 15-sekundnem vzorcu. Da bi se ugotovilo ali je težava v pridobitvi ali prepoznavanju lika, je bilo tudi prešteto v koliko okvirjev je obraz najden (v tabeli 8.1: Pridobljeni obrazi) in v koliko je lik prepoznan, ki naj bi bil zamenjan (v tabeli 8.1: Prepoznani obrazi). Glede na to, da je bil obraz pridobljen v vsakem okvirju, sklepamo da je težava v prepoznavanju obraza.

Poskusili smo odpraviti napako, tako, da smo zmanjšali prag za prepoznavanje obrazov, vendar je to povzročilo zamenjavo nepravilnih obrazov v okvirju. Primere pravilne zamenjave, ko so liki zaznani iz strani in je prag zmanjšan prikazuje slika 8.4. Ker je bil cilj programa zamenjava specifičnega obraza, tega nismo pustili v končni različici programa.



Slika 8.3: Pravilna zamenjava obraza na praktičnih primerih frontalnih obrazov likov.



Slika 8.4: Pravilna zamenjava obraza na praktičnih primerih zaznanih likov iz strani.

Poglavlje 9

Zaključek

V diplomski nalogi smo prikazali, kaj je tehnologija globokih ponaredkov in kaj so to globoki ponaredki. Preučili smo zgodovino razvoja globokih ponaredkov in različnih oblik, v katerih so se pojavljali skozi leta. Prav tako so bile kategorizirane vse vrste globokih ponaredkov, ki obstajajo danes, hkrati pa so bili podani primeri za vsako kategorijo.

Poleg tega so bile opisane še najbolj pogoste in popularne uporabe in zlorabe tehnologij globokih ponaredkov. Za vsak primer je bilo podanih veliko primerov ter analiziran vpliv, ki ga imajo nekateri primeri na resnični svet.

Nato smo naredili pregled bolj aktualnih in popularnih metod in pristopov, ki se uporablajo v tehnologiji globokih ponaredkov. Predstavili smo še ostalo programsko opremo, ki se ukvarja z ustvarjanjem globokih ponaredkov. Poleg tega je bila podrobno prikazana metoda, na kateri je temeljil program za zamenjavo obraza, ki je bil narejen v okviru te diplomske naloge.

Na koncu je bil predstavljen še program za zamenjavo obraza, ki je bil narejen kot rezultat te diplomske naloge. Namenjen je uporabi v izobraževalnem okolju. Z njim bi si uporabnik lahko ogledal sebe, kako igra v izbranem filmu. Ta program zadostuje za zamenjavo obraza osebe v videu s poljubnim

obrazom v relativno hitrem času. Trenutno program ne nudi uporabniku prijaznega vmesnika. Z implementacijo tega manjkajočega dela pa upamo, da bo program uporaben za izobraževalne cilje. Program se bi s pomočjo optimizacije izvajal v skoraj realnem času. To bi dosegli tako, da se bi začetni okvirji shranjevali v medpomnilnik (angl. buffer) in se bi video skupaj z audio lahko predvajal med obdelavo ostalih okvirjev.

Literatura

- [1] Insightface. URL <https://insightface.ai/>.
- [2] Deep learning "vae-gan". URL https://blog.csdn.net/qq_29367075/article/details/110849112.
- [3] Neta Alexander. Obsolescence, forgotten: "survivor holograms", virtual reality, and the future of holocaust commemoration. *Cinergie–Il Cinema e le altre Arti*, (19):57–68, 2021.
- [4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [5] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 353–360, 1997.
- [6] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 353–360, 1997.
- [7] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros. Everybody dance now. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5933–5942, 2019.

- [8] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Sims-wap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2003–2011, 2020.
- [9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.
- [10] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5203–5212, 2020.
- [11] Faceswap. Faceswap. URL <https://faceswap.dev/>.
- [12] Pisana Ferrari. Deepfake lip syncing technology could help translate film and tv without losing an actor’s original performance. URL <https://www.capstan.be/deepfake-lip-syncing-technology-could-help-translate-film-and-tv-without-losing-an-actors-original-performance/>.
- [13] Gabriela Galindo. Xr belgium posts deepfake of belgian premier linking covid-19 with climate crisis. *The Brussels Times*, 14, 2020.
- [14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [15] Drew Harwell. Scarlett johansson on fake ai-generated sex videos:‘nothing can stop someone from cutting and pasting my image’. *Washington Post*, 31:12, 2018.

- [16] Iperov. Deepfacelive. URL <https://github.com/iperov/DeepFaceLive>.
- [17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [18] Charlotte Jee. An indian politician is using deepfake technology to win new voters, 2020.
- [19] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.
- [20] Dami Lee. Deepfake salvador dalí takes selfies with museum visitors. *The Verge*, 2019.
- [21] Jae Hyun Lim and Jong Chul Ye. Geometric gan. *arXiv preprint arXiv:1705.02894*, 2017.
- [22] Bahar Uddin Mahmud and Afsana Sharmin. Deep insights of deepfake technology: A review. *arXiv preprint arXiv:2105.00192*, 2021.
- [23] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, 54(1):1–41, 2021.
- [24] Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. {CT-GAN}: Malicious tampering of 3d medical imagery using deep learning. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 461–478, 2019.
- [25] Illinois Holocaust Museum and Education Center. Interactive holograms: Survivor stories experience. URL <https://www.ilholocaustmuseum.org/exhibitions/survivor-stories-experience/>.

- [26] Inc. d/b/a Reface Neocortex. Reface. URL <https://hey.reface.ai/>.
- [27] Scripps News. At the illinois holocaust museum, holograms forever preserve survivors. URL <https://www.youtube.com/watch?v=1jVxkFSy2Fk>.
- [28] O Oakes. „deepfake“ voice tech used for good in david backham malaria campaign. *PR Week*, 2019.
- [29] pawangfg. Variational autoencoders. URL <https://www.geeksforgeeks.org/variational-autoencoders/>.
- [30] Rahul Roy. Generative adversarial network (gan). URL <https://www.geeksforgeeks.org/generative-adversarial-network-gan/>.
- [31] Raphael Satter. Deepfake used to attack activist couple shows new disinformation frontier. *Reuters, July*, 2020.
- [32] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. *Advances in Neural Information Processing Systems*, 32, 2019.
- [33] Supasorn Suwajanakorn. Synthesizing obama: Learning lip sync from audio. URL <https://www.youtube.com/watch?v=9Yq67CjDqvw>.
- [34] Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: learning lip sync from audio. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017.
- [35] Patrick Ryan The Sto. The dark sto: Blackmail inflation w/ patrick ryan. URL <https://www.youtube.com/watch?v=x1mh9HYqc>.
- [36] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2387–2395, 2016.

- [37] James Vincent. New ai deepfake app creates nude images of women in seconds. *The Verge*, 27, 2019.
- [38] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.
- [39] Olivia Wiles, A Koepke, and Andrew Zisserman. X2face: A network for controlling face generation using images, audio, and pose codes. In *Proceedings of the European conference on computer vision (ECCV)*, pages 670–686, 2018.
- [40] Erin Winick. Actors are digitally preserving themselves to continue their careers beyond the grave, 2018.
- [41] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lemitsky. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9459–9468, 2019.