



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INTELIGENTNÍCH SYSTÉMŮ
DEPARTMENT OF INTELLIGENT SYSTEMS

ANTROPOMETRICKÁ EVALUACE GENEROVANÝCH SNÍMKŮ OBLÍČEJE

ANTHROPOMETRIC EVALUATION OF GENERATED FACE IMAGES

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

JAKUB MIKYŠEK

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. TOMÁŠ GOLDMANN

BRNO 2024

Zadání bakalářské práce



Ústav: Ústav inteligentních systémů (UITS) 154467
Student: **Mikyšek Jakub**
Program: Informační technologie
Název: **Antropometrická evaluace generovaných snímků obličeje**
Kategorie: Umělá inteligence
Akademický rok: 2023/24

Zadání:

1. Seznamte se state-of-the-art neuronovými sítěmi pro generování syntetických snímků obličeje.
2. Nastudujte jaké antropometrické proporce jsou významné pro určování identity osob podle obličeje.
3. Navrhnete řešení pro analýzu antropometrických proporcí mezi generovanými a reálnými snímkami obličeje.
4. Navržené řešení implementujte v programovacím jazyce Python.
5. Proveďte vyhodnocení vašeho řešení alespoň na 50 dvojicích generovaných a reálných snímků.

Literatura:

- CHOI, Yunjey, et al. Stargan v2: Diverse image synthesis for multiple domains. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020. p. 8188-8197.
- GUO, Jianzhu, et al. Towards fast, accurate and stable 3d dense face alignment. In: *European Conference on Computer Vision*. Cham: Springer International Publishing, 2020. p. 152-168.

Při obhajobě semestrální části projektu je požadováno:

Body 1 a 2.

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Goldmann Tomáš, Ing.**

Vedoucí ústavu: Hanáček Petr, doc. Dr. Ing.

Datum zadání: 1.11.2023

Termín pro odevzdání: 9.5.2024

Datum schválení: 10.2.2024

Abstrakt

Tato práce se zabývá porovnáním uměle vygenerovaných obličejů s reálnými snímky pomocí analýzy klíčových bodů obličeje a měřením proporcí mezi těmito body. Zkoumá také oblasti týkající se umělých neuronových sítí, přičemž se zaměřuje na síť GAN, které dokážou generovat uměle vytvořené snímky obličeje. Zkoumá jejich proces, architekturu a dostupné modely. Cílem práce je zhodnotit, jak se uměle vygenerované snímky liší od reálných, a zjistit, v jakých proporcích je rozdíl největší.

Abstract

This work is focused on comparing artificially generated faces with real images by analyzing key facial landmarks and measuring the proportions between these landmarks. It also explores areas related to artificial neural networks, focusing on GANs that can generate artificially generated facial images. It explores their process, architecture and available models. The aim of the work is to evaluate how artificially generated images differ from real ones, and to find out in which proportions the difference is the largest.

Klíčová slova

antropometrie, antropometrické body, rozpoznání obličeje, neuronové sítě, GAN, generované snímky obličeje

Keywords

anthropometry, anthropometric landmarks, face recognition, neural networks, GAN, generated face images

Citace

MIKYŠEK, Jakub. *Antropometrická evaluace generovaných snímků obličeje*. Brno, 2024. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Tomáš Goldmann

Antropometrická evaluace generovaných snímků obličeje

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Tomáše Goldmanna. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

Jakub Mikyšek
7. května 2024

Poděkování

Děkuji svému vedoucímu panu Ing. Tomáši Goldmannovi za poskytnutí rad ohledně obsahu práce a za možnost osobních konzultací k probrání dotazů.

Obsah

1	Úvod	4
2	Antropometrie lidské tváře	5
2.1	Historie	5
2.2	Antropometrické body a proporce	7
3	Neuronové sítě pro generování syntetických snímků obličeje	11
3.1	Úvod do neuronových sítí	11
3.2	Generativní neuronové sítě	14
3.3	Generování syntetických snímků obličeje umělou inteligencí	16
3.4	Detekce antropometrických bodů v obrazu	24
3.5	Rozpoznávání a ověřování obličeje	27
4	Návrh řešení a implementace analýzy antropometrických proporcí	30
4.1	Blokové schéma návrhu	31
4.2	Datová sada reálných snímků obličeje	31
4.3	Generování umělých snímků obličeje	32
4.4	Informace o antropometrických bodech na obličeji	34
4.5	Antropometrická analýza významných bodů obličeje	36
4.6	Porovnání uměle vygenerovaných a reálných snímků obličeje	37
4.7	Pomocné skripty a funkce	38
5	Experimenty a vyhodnocení	40
5.1	Antropometrická analýza	40
5.2	Srovnávací analýza	42
5.3	Kombinovaná analýza	44
5.4	Obecný pohled na dvojice	46
6	Závěr	48
	Literatura	49

Seznam obrázků

2.1	Fairbankův obličejobý měřicí přístroj, odborný nástroj pro měření lidského obličeje [28].	6
2.2	Antropometrické body na obličeji [5].	7
3.1	Model procesu perceptronu [32].	12
3.2	<i>Deep learning</i> neuronová síť s několika skrytými vrstvami [33].	13
3.3	GAN diagram obsahující model generátoru, generující výsledky z dat získaných na vstupu a diskriminační model, určující věrohodnost získaných vstupů od generátoru a datové sady. Postavené na bázi hry pro 2 – minimax [10].	15
3.4	Struktura modelu ELEGANT pro změnu atributů a následnou syntézu snímků obličeje [7].	17
3.5	Přehled sítě StarGAN [2].	18
3.6	Výsledky STGAN pro úpravu atributů obličeje [23].	19
3.7	Struktura modelu SGGAN – a) generátor G, b) diskriminátor D, c) segmentátor S, během tréninku S poskytuje G prostorové pokyny, aby se zajistilo, že generované obrazy odpovídají vstupním segmentacím, d) tréninková procedura zajišťující realistický výstup [8].	20
3.8	Celkový tréninkový proces u MaskGAN. Úprava chování simulovaného tréninku lze rozdělit do dvou fází. Po načtení předtrénovaného modelu husté mapovací sítě a MaskVAE tyto dvě fáze iterativně aktualizujeme, dokud model nekonverguje [22].	21
3.9	Na obrázku je možné vidět porovnání tradičního generátoru (a) a stylově založeného generátoru (b), kterým je StyleGAN [17].	22
3.10	Dvě sady snímků byly vygenerovány z odpovídajících latentních kódů (<i>Source A</i> a <i>Source B</i>). Zbytek obrázků byl vygenerován pomocí stylu z zdroje B a následně zdroje A. Kopírování stylů odpovídající hrubému rozlišení ($4^2 - 8^2$), umožňuje přenést vlastnosti vyšší perspektivy, jako jsou póza, styl účesu, tvar obličeje a brýle, ze zdroje B a styl barev (oči, vlasy, nasvícení) je přenesen ze zdroje A [17].	23
3.11	Regresivní kaskáda pro určování odhadovaného rozložení antropometrických bodů na obličeji v jednotlivých stádiích. Již po první úrovni kaskády se chyba výrazně sníží [19].	25
3.12	68 (x, y) souřadnic pro poskytnutí anotace datovou sadou iBUG 300-W [30].	26
3.13	Ukázka různých přístupů modelů využitelných v rámci Deepface k detekování lidského obličeje [31].	28
3.14	Příklad rozdílu mezi ztrátovou funkcí (<i>loss function</i>) Softmax a ArcFace na 8 různých identitách. Na základě normalizace rysů jsou všechny rysy obličeje posunuty do obloukového prostoru (<i>the arc space</i>) s pevným poloměrem [6].	29

4.1	Blokové schéma návrhu práce.	31
4.2	Část snímků podmnožiny datové sady CelebAMask-HQ.	32
4.3	Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3.	33
4.4	Ukázka vygenerovaných snímků obličeje pomocí modelu StyleGAN3.	34
4.5	Vybrané zkoumané antropometrické body na obličeji. Názvy bodů jsou konstanty s hodnotou odpovídající indexu na masce modelu MediaPipe.	35
4.6	Ukázka vizualizace výsledků detekce antropometrických bodů obličeje modelu MediaPipe.	36
5.1	Kruhový diagram pro zobrazení úspěšnosti porovnání obličejů rámcem Deepface. Zelená část značí shodu v porovnání (<i>Success</i>) a červená značí neúspěch (<i>Failure</i>).	42
5.2	Korelační diagram pro zobrazení vzdálenosti vektorů dvojic snímků obličejů. Červená přerušovaná čára zastupuje prahovou hodnotu (<i>threshold</i>), rozdoující o úspěšnosti porovnání.	43
5.3	Procentuální zastoupení na intervalech pro všechny naměřené vzdálenosti.	44
5.4	Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3 s vysokou hodnotou vzdálenosti mezi vektory obličeje.	46
5.5	Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3 s nízkou hodnotou vzdálenosti mezi vektory obličeje.	47

Kapitola 1

Úvod

Lidský obličej a jeho jedinečnost od obličejů dalších osob nám umožňuje ho využívat jako nástroj. Například pro odemykání chytrého telefonu pomocí rozpoznání obličeje. Vyhodnocování věrohodnosti snímků obličeje je zásadní hlavně z pohledu forenzního zkoumání a soukromí fyzických osob. Prostředkem pro získání vědomostí o obličeji jedince jsou antropometrické body na obličeji, tvořící významné proporce pro měření obličeje a jeho celkovou analýzu. Získaná měření mohou být využita v mnoha oblastech, jako je medicína, antropologie nebo kriminalistika. Kromě obecných dat je důležité sbírat také specifická data pro konkrétní etnické skupiny, což umožní přesnější zkoumání a kvalitnější zásahy v různých oblastech výzkumu a praxe. Tento přístup podporuje větší diverzitu a inkluzi ve výzkumu lidských obličejů a souvisejících oborů.

Rostoucí zájem o umělou inteligenci a neuronové sítě posouvá toto odvětví raketovou rychlosťí kupředu. Algoritmus, který tu byl před rokem může být následující měsíc nahrazen novým a lepším algoritmem, který problematiku řeší ještě lépe. Práce se zaměřuje primárně na síť GAN, které jsou schopny generovat vizuální obsah za pomocí umělých neuronových sítí. Dnes je možné generovat mimo snímky i videa a další multimediální obsah. Konkrétně bude pohlíženo na rámce umožňující generovat umělé snímky obličeje, včetně náhledu na jejich strukturu a zástupce.

Další klíčovou oblastí výzkumu neuronových sítí jsou modely pro rozpoznávání obličeje. Tyto modely jsou navrženy tak, aby dokázaly detektovat a identifikovat tváře jednotlivců v různých kontextech a prostředích, což zahrnuje různé úhly pohledu, osvětlení a výrazy obličeje. Nejen, že tyto modely umožňují porovnání dvojcí obličejů pro zjištění podobnosti nebo odlišnosti, ale také pomáhají identifikovat jedince na základě specifických rysů obličeje. V této práci jsou modely využity pro detekování antropometrických bodů na obličeji k následné analýze obličeje a ke rozpoznání obličeje pro srovnání podobnosti dvou obličejů, včetně vyhodnocení na základě výpočtu vzdálenosti vektorů obličeje.

Momentálně je nedostatek výzkumů týkajících se podobnosti uměle vygenerovaných snímků obličeje s reálnými snímky, což ztěžuje hodnocení pokroku v oblasti generování umělých snímků pomocí GAN sítí. Cílem této práce je poskytnout tato data a vyhodnotit rozdíly a podobnosti mezi dvojicemi snímků obličeje, čímž přispěje k dalšímu rozvoji této oblasti.

Kapitola 2

Antropometrie lidské tváře

Jedinečnost každé tváře je ovlivněná antropometrickými rysy, které jsou determinovány genetickými a socioekonomickými faktory [21]. Antropometrie je biologická věda, která se zabývá studiem a měřením proporcí částí lidského těla. Tento pojem je odvozen z řeckých slov *antropos*, což znamená člověk, a *metreo*, což znamená měřit. Její použití můžeme vidět v odvětvích jako je antropologie a kriminalistika, ale také v lékařství, sportu či oděvním průmyslu. Díky vývoji technologií v oblasti počítacového vidění a neuronových sítí, existuje celá řada nástrojů a algoritmů pro detekci lidského obličeje a antropometrických bodů pro identifikaci člověka [14].

Antropometrická analýza je zvlášť významná v oblasti medicíny, kde kvantitativní porovnání antropometrických údajů s mírami pacientů před a po operaci napomáhá plánování a hodnocení plastických a rekonstrukčních operací. V oblasti forenzní antropologie jsou antropometrické údaje klíčové při odhadu vzhledu jedinců na základě jejich ostatků [5].

Pojem také důležitý k definování je biometrie. Biometrie je statistický a analytický přístup k zkoumání měřitelných biologických charakteristik živých organismů. Jeho cílem je praktické využití při jednoznačné identifikaci nebo verifikaci jednotlivců [4]. Tento přístup, společně s antropometrickou analýzou, poskytuje komplexní nástroje pro porozumění a využití individuálních charakteristik lidského těla v různých oborech.

2.1 Historie

Pro pochopení významu antropometrie je žádoucí se podívat do historie celé problematiky, na její vznik a vývoj. Jak na ni bylo nahlíženo v době jejího počátku, jak se liší praktikování dnes a co za nástroje je pro měření potřeba.

Zkoumání lidských rysů v historii

Pojem antropometrie poprvé zavedl francouzský badatel Alphonse Bertillon [14] (nar. 1853) a to jako, systém identifikací závislých na neméněm charakteru určitých rozměrů částí lidského těla. Došel k tomu na základě experimentu, kde zjistil, že některé tělesné rysy se v dospělosti téměř nemění a z toho vyvodil, že lze jednoznačně identifikovat každého člověka, protože každý jednotlivec je dokonale odlišitelný od ostatních.

Tento systém se velice rychle osvědčil v policejních metodách pro stanovení totožnosti pachatele. Svoji popularitu získal v roce 1883 ve Francii a následně se rozšířila ve výkonu spravedlnosti ve většině civilizovaných zemí, s tím se mimo jiné pojí i částečné přijmutí systému identifikace na základě papilárních linií na vnitřních straně článků prstů u osob.

Bertillon vybral podle svého názoru pět klíčových měr pro základ jeho systému [14]: délku a šířku hlavy, délku prostředníčku, délku levé nohy a délku loktu (měřenou od předloktí po konec prostředníčku).

Pro zvýšení přesnosti byla rovněž zaznamenána výška jedince, délka jeho malíčku a barva očí. Tyto všechny informace umožňovaly vyhledat hledané osoby. Avšak tento proces byl zdlouhavý, docházelo k chybám při zapisování údajů, bylo potřebné specifické vybavení a vyškolený pracovník. I proto se nakonec od této metody upustilo a plně se přešlo na systém identifikace podle otisku prstu [14].

Mezi faktory ovlivňující antropometrii člověka bylo doposud zkoumáno pouze genetické hledisko, avšak existuje alespoň ještě jeden aspekt. Je proto nezbytné nahlédnout do počátků studia lidské fyzické postavy – v roce 1829, kdy Villermé, a 1831, kdy Quetelet, poprvé přinesli zmínku o této problematice z hlediska metriky výšky populace.

Ve pozdějším zkoumání francouzských historiků tradice Annales v 60. letech 20. století, odhalili souvislost mezi socioekonomickými faktory a lidskou výškou. Jejich výzkum naznačuje, že biologický vývoj jedince není ovlivněn pouze genetickými faktory, ale také **socioekonomickým prostředím**, včetně nutriční kvality jídla [21].

Vybavení pro měření lidského obličeje

Pro precizní vyhodnocení lidského obličeje, je třeba ho nejprve řádně změřit. Variant je zde relativně dost – je možné si vystačit s obyčejným pravítkem a úhloměrem, ale ta budou poskytovat pouze omezenou přesnost, proto existují odbornější alternativy, k těm používanějším patří Fairbankův obličejevý měřicí přístroj (*Fairbanks Facial Measurement Gauge*), umožňující měřit až 14 různých oblastí na obličeji a to již s velkou přesností, přesto pořád může dojít k drobným odchylkám.



Obrázek 2.1: Fairbankův obličejevý měřicí přístroj, odborný nástroj pro měření lidského obličeje [28].

Mezi další odborné měřící přístroje používané v lékařství patří kraniometr nebo kefalometr. Zmíněné pomůcky fungují na bázi kruhovitých otevíracích klešťí pro měření hlavy, v různých stádiích života jedince.

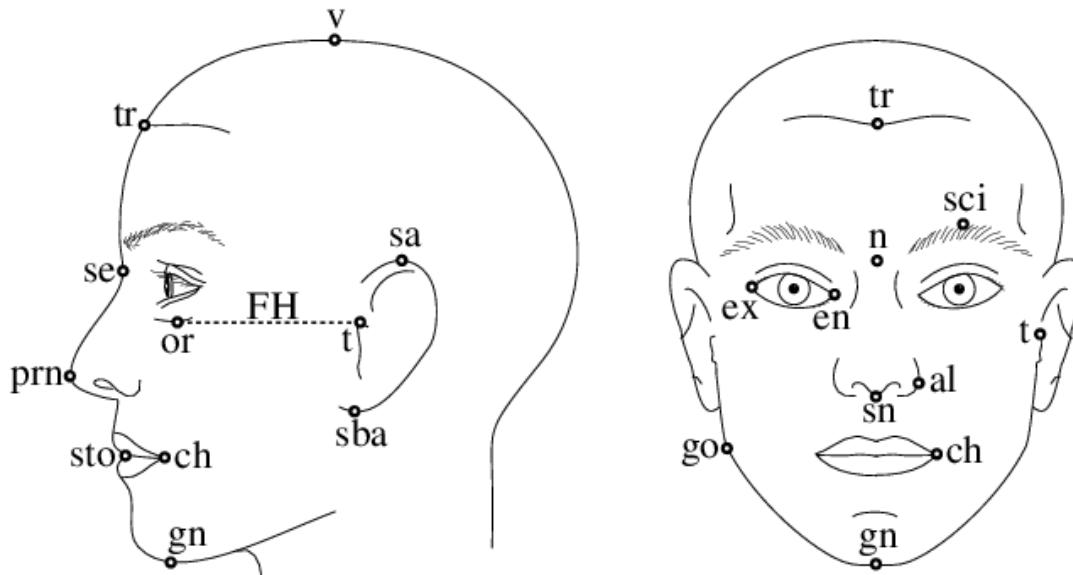
Zaznamenávání údajů o lidském těle nám pomáhá řešit problémy (kriminalistika, antropologie, lékařství), rozvíjet poznatky o člověku a vynalézat nové použití pro fakt jedinečnosti lidské tělesné schránky (odemykání chytrého telefonu pomocí rozpoznání obličeje), je tedy důležité se v této oblasti i nadále rozvíjet a zkoumat ji.

2.2 Antropometrické body a proporce

Využití antropometrických měření je pro chirurgy zásadní při plánování operací obličeje, ať už se jedná o rekonstrukční operace po úrazech a onkologických resekciích, nebo o estetické zákroky. Nicméně použití jednoho standardu antropometrických rysů pro všechny etnické skupiny není vhodné, neboť může vést k esteticky neuspokojivým výsledkům. Proto je důležité, aby pro každou etnickou skupinu byly k dispozici specifické obličejové normy. Toho se snažil profesor Leslie Gabriel Farkas, který je považován za průkopníka moderní antropometrie obličeje. On a jeho spolupracovníci zkoumali v multicentrické studii kraniofaciální charakteristiky na 25 etnických skupinách a vyvinuli měření, které obsahovalo 14 parametrů pro každou z těchto etnik [36].

Antropometrické body (*landmarks*) jsou konkrétní místa na zkoumaném subjektu, určených podle viditelných nebo hmatných znaků (kůže nebo kostí) [5].

S cílem rozšířit znalosti etnické antropometrie v řecké populaci byla v roce 2015 provedena série antropometrických analýz za účelem získání významných bodů obličeje. Studie zahrnovala 152 dobrovolníků, z nichž 78 byli muži a 74 ženy ve věkovém rozmezí 18 až 30 let. Celkem bylo provedeno 31 antropometrických měření, z toho 25 jednotlivých a 3 párová. Před samotným měřením byly na obličeji každého jednotlivce vyznačeny povrchové antropometrické body z tabulky 2.1. Každé měření bylo provedeno dvakrát stejným vyšetřovatelem a následně byla použita vypočtená průměrná hodnota [36].



Obrázek 2.2: Antropometrické body na obličeji [5].

<i>Glabella</i> (g)	Nejvýraznější středový bod mezi obočím
<i>Opisthocranion</i> (op)	Nejvzdálenější bod od bodu glabella na zadní části hlavy
<i>Trichion</i> (tr)	Bod na linii vlasů ve střední části čela
<i>Nasion</i> (n)	Bod ve střední linii kořene nosu
<i>Subnasale</i> (sn)	Střední bod, kde se setkává dolní okraj nosní přepážky a povrch horního rtu
<i>Gnathion</i> (gn)	Nejnižší bod ve střední linii na dolním okraji dolní čelisti
<i>Zygion</i> (zy)	Nejbočnejší bod každého jařmového oblouku
<i>Gonion</i> (go)	Vrchol úhlu dolní čelisti
<i>Stomion</i> (sto)	Pomyslný bod na křížení vertikální střední čáry obličeje a horizontální štěrbiny mezi jemně zavřenými rty, se zavřenými zuby v přirozené poloze
<i>Endocanthion</i> (en)	Bod na vnitřním okraji oční štěrbiny
<i>Exocanthion</i> (ex)	Bod na vnějším okraji oční štěrbiny
<i>Alare</i> (al)	Nejbočnejší bod na každém okraji nosního křídla
<i>Cheilion</i> (ch)	Bod umístěný na každém spojení rtů
<i>Labiale superius</i> (ls)	Střední bod horní linie vermillionu
<i>Labiale inferius</i> (li)	Střední bod spodní linie vermillionu
<i>Superaurale</i> (sa)	Nejvyšší bod na volném okraji ušního boltce
<i>Subaurale</i> (sba)	Nejnižší bod na volném okraji ušního lalůčku
<i>Tragion</i> (t)	Bod na horním okraji zevního zvukovodu

Tabulka 2.1: Analyzované antropometrické body [36].

Provedená měření [36] lze s pomocí tabulky 2.1 a obrázku 2.2 popsat takto:

- Měření hlavy
 - horizontální obvod hlavy (g-op-g),
 - výška čela (tr-g),
 - rozšířená výška čela (tr-n).
- Měření obličeje
 - oblouk horní čelisti (t-sn-t),
 - čelistní oblouk (t-gn-t),
 - šířka obličeje (zy-zy),
 - šířka dolní čelisti (go-go),
 - spodní výška obličeje (sn-gn),

- fyziognomická výška obličeje (tr-gn),
 - morfologická výška obličeje (n-gn),
 - fyziognomická výška horní části obličeje (n-sto),
 - výška spodní třetiny obličeje (sto-gn),
 - speciální výška obličeje (en-gn),
 - speciální výška horního obličeje (g-sn).
- Měření orbitů
 - binokulární průměr (ex-ex),
 - interkantální vzdálenost (en-en),
 - délka oční štěrbiny (ex-en).
 - Měření nosu
 - šířka nosu (al-al),
 - výška nosu (n-sn),
 - sklon nosního můstku (sklon nosu, ni).
 - Měření labio-orální oblasti
 - šířka úst (ch-ch),
 - mediální výška vermillionu horního rtu (ls-sto),
 - mediální výška vermillionu dolního rtu (sto-li),
 - mediální výška kůže horního rtu (sn-ls),
 - mediální vertikální délka horního rtu (sn-sto).
 - Měření uší
 - délka ucha (sa-sba),
 - sklon mediální podélné osy ušního boltce (sklon ucha, ei).

Po zpracování výsledků byla provedena jejich vzájemná komparace. Při porovnání mužských a ženských výsledků byly zjištěny některé významné rozdíly. Zejména výška čela (tr-g) byla zaznamenána jako extrémně menší u žen. Další rozdíly byly pozorovány v délce nosu a uší, šířce úst, velikosti a vzdálenosti očí. V těchto ohledech byly zaznamenány větší rozměry u mužů, avšak rozdíly nebyly tak výrazné jako v případě výšky čela. Horní rty byly mírně větší u řeckých žen, zatímco dolní rty byly větší u řeckých mužů [36].

Další analýzou pro měření obličeje je celkový index obličeje [15] (*total facial index, TFI*). Jedná se o poměr morfologické výšky obličeje (získaný měřením vzdálenosti n-gn) a maximální šířky obličeje (vzdálenost mezi dvěma jařmovými oblouky zy-zy). Naměřená hodnota TFI slouží k určení typu obličeje podle stupnice Martin-Saller a lze jej vypočítat podle vzorce:

$$TFI = \frac{n - gn \times 100}{zy - zy} \quad (2.1)$$

Jednotlivé výsledky měření TFI jsou použity k určení výskytu určitých typů obličeje, rozdělené podle rozmezí hodnot umístěných v tabulce 2.2. Rozmezí jsou určené na základě škály Martin-Saller [15].

Název	Rozmezí
hypereuryprosopic	$TFI \leq 78,9$
curyprosopic	$79,0 < TFI < 83,9$
mesoprosopic	$84,0 < TFI < 87,9$
leptoprosopic	$88,0 < TFI < 92,9$
hyperleptoprosopic	$TFI \geq 93,0$

Tabulka 2.2: Jednotlivé typy obličeje na základě hodnoty TFI [15].

Mimo zmíněný celkový index obličeje, existují i jiné typy poměrů obličeje, které zohledňují další proporce lidské tváře. Příkladem je poměr pro horní část obličeje.

Významné antropometrické body na lidském obličeji jsou klíčové pro přesné měření a analýzu obličeje, což je důležité pro mnoho aplikací. Tyto body mohou sloužit jako referenční body pro určení geometrických vlastností obličeje a umožnit tak vytvoření datových sad pro trénování algoritmů strojového učení, stejně jako poskytují cenný zdroj pro chirurgii specializující se na operace obličeje. Takové datové sady jsou základem pro vývoj a zdokonalování technologií spojených s obličejem, které naleznou využití například v biometrii, zábavním průmyslu, medicíně nebo bezpečnostních aplikacích.

Kapitola 3

Neuronové sítě pro generování syntetických snímků obličeje

Neuronové sítě (angl. *neural networks*), známé také jako **umělé neuronové sítě** [3] (angl. *artificial neural networks*), jsou podmnožinou strojového učení (angl. *machine learning*) – procesu trénování počítačových programů a jsou základem algoritmů hlubokého učení (angl. *deep learning algorithms*). Jejich název a struktura jsou inspirovány lidským mozkem, protože napodobují způsob, jak si biologické neurony mezi sebou předávají signály. Umělé neuronové sítě se rozvíjejí v různých směrech, nabízející tak specifická využití v konkrétních oblastech. Mezi ně patří i generativní neuronové sítě, které jsou určeny pro tvorbu nového obsahu nebo modifikaci již existujícího.

Neuronové sítě jsou prostředkem **strojového učení**, ve kterém se počítač učí provádět určitý úkol na základě analýzy trénovacích případů. Příklady jsou obvykle ručně označeny – například vložím obrázek do rozpoznávacího systému a označím, co se na obrázku nachází, to stejně mohu udělat s celou sadou obrázků.

Na základě uživatelem dopředu vytvořených označení, je umělá inteligence schopná se učit, které výsledky provedla správně a vytvoří si v případě detekce objektu – vizuální vzory, které konzistentně korelují s určitými označeními a bude se jimi řídit dále (například u obrázků auta detekuje, že častým jevem jsou kola, proto se na tuhle oblast začne soustředit a dojde k přesnějším výsledkům). Čím více takových dat, tím přesnější se umělá inteligence může stát [12].

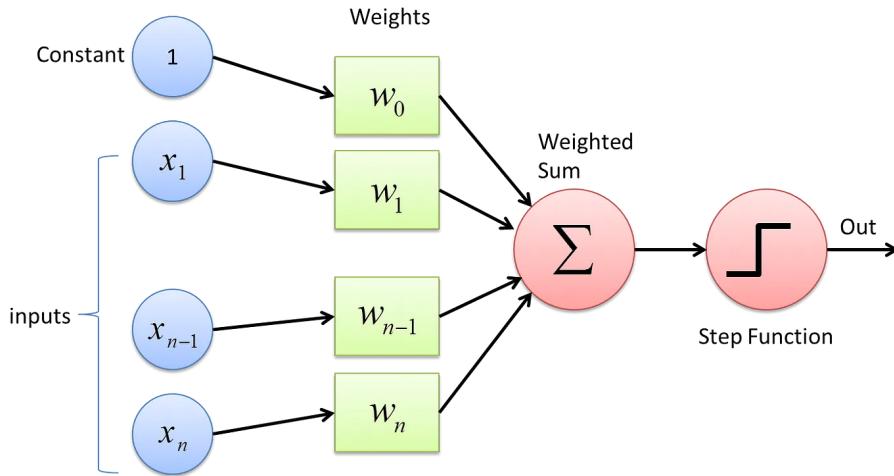
Umělá inteligence zaměřená pouze na jeden úkol se nazývá *artificial narrow intelligence*, opakem pak je *artificial general intelligence*, které nemá konkrétní zaměření – na jejím vývoji se momentálně pracuje.

3.1 Úvod do neuronových sítí

Neuronové sítě se skládají z uzlových vrstev, které obsahují vstupní vrstvu (angl. *input layer*), jednu nebo více skrytých vrstev (angl. *hidden layers*) a výstupní vrstvy (angl. *output layer*). Každý uzel se připojuje k jinému uzlu a má přiřazené váhy (angl. *weight*) a práh (angl. *threshold*). Pokud je hodnota některého z jednotlivých výstupu vyšší než stanovená prahová hodnota, aktivuje se uzel a předávají se data do následující vrstvy. V případě, že toto neplatí, nedochází k dalšímu předávání dat [3].

Perceptron [32] je jednovrstvá neuronová síť. Vícevrstvý perceptron se označuje jako neuronové sítě (*neural networks*), ty ve své podstatě fungují stejně jako perceptron. Sestává z 4 klíčových částí:

1. vstupní hodnoty (x_n),
2. váhy ($w_0 \dots w_n$),
3. součet násobení jednotlivých vstupů se svými váhami (\sum),
4. aktivační funkce.



Obrázek 3.1: Model procesu perceptronu [32].

Všechny vstupy (x_n) jsou vynásobeny s jejich váhami (w_n), výsledek operace $x_n w_n$ může být nazýván například k . Nad těmito výsledky k je následně provedena suma, ke které je možné přičíst i nějakou odchylku (bias b):

$$sum = \sum_{i=0}^n k_i + b$$

Získaná hodnota sum je aplikována na aktivační funkci, která obsahuje určitý práh, určující výstup v souladu s typem zvolené aktivační funkce (např. lineární, skoková, sigmoidní, ReLU, ...).

Ztrátová funkce [11] (*loss function*) je nástrojem pro hodnocení účinnosti algoritmů strojového učení ve vztahu k datové sadě. Konkrétně měří, jak přesně algoritmus dosahuje zamýšlených výsledků a vyjadřuje, jak úspěšně představuje daný výsledek. Čím nižší je hodnota ztrátové funkce, tím je model přesnejší, protože naznačuje menší rozdíl mezi očekávanými a předpovězenými výsledky. V průběhu trénování se váhy a parametry modelu mění tak, aby byla minimalizována hodnota ztrátové funkce.

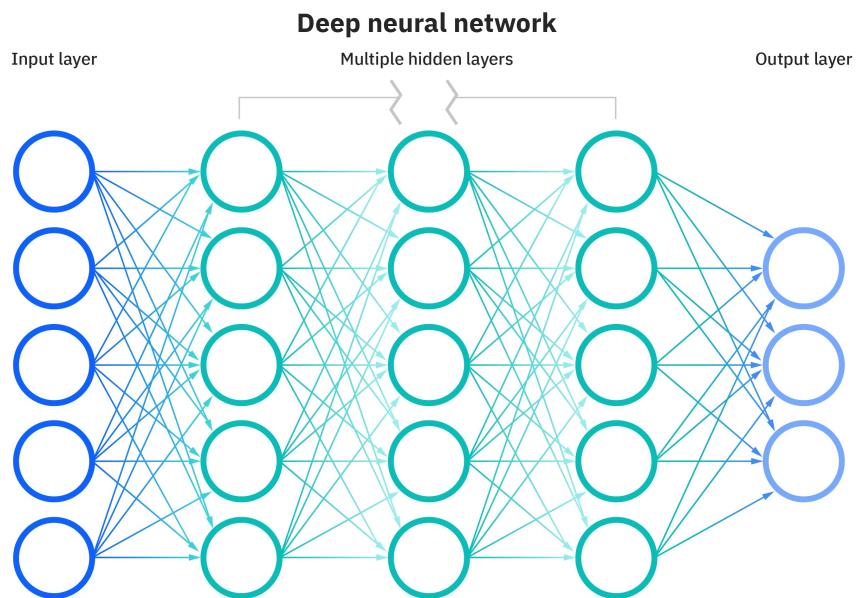
Existuje mnoho typů ztrátových funkcí, z nichž každá je vhodná pro určité úlohy strojového učení. Pro klasifikační úlohy často používaná křížová entropie (*cross-entropy*), která měří rozdíl mezi pravdivými a předpovězenými pravděpodobnostmi tříd. V regresních úlohách, kde se předpovídají kontinuální hodnoty, se často uplatňuje střední kvadratická chyba (*Mean Squared Error*) [11].

Třívrstvá neuronová síť představuje pokročilý model umělé inteligence, který vychází z konceptu perceptronu, avšak zahrnuje více vrstev pro komplexnější zpracování informací. Tato síť se skládá ze tří hlavních vrstev: vstupní vrstvy, skryté vrstvy a výstupní vrstvy.

Zpětné šíření (*backpropagation*) je klíčovým mechanismem učení neuronových sítí. Po prezentaci vstupních dat a generování odpovědi výstupní vrstvou dochází k porovnání této odpovědi s očekávanými výstupy. Chybový signál se pak šíří zpětně skrz síť a upravuje váhy mezi neurony, aby se minimalizovala chyba predikce. Tímto způsobem síť postupně optimalizuje své váhy a dosahuje lepší schopnosti generování nových dat. Opakem *backpropagation* je *feedforward* šíření, které postupuje jedním směrem od vstupu do výstupu [3].

Dalším zásadním pojmem je hluboké učení (*deep learning*). Rozdíl mezi *deep learning* a neuronovými sítěmi nemusí být ale vždy jasný. Oba termíny spolu souvisí a mohou se prolínat – „hloubka“ v hlubokém učení značí hloubku vrstev v neuronové síti, tedy to jak moc je neuronová síť rozsáhlá.

Neuronová síť, která je složená **z více než tří vrstev** (včetně vstupní a výstupní vrstvy), **může být brána jako algoritmus hlubokého učení** (angl. *deep learning algorithm*), v opačném případě (3 nebo méně vrstev, včetně) se jedná jen o základní neuronovou síť [3].



Obrázek 3.2: *Deep learning* neuronová síť s několika skrytými vrstvami [33].

Konvoluční neuronové sítě

Pojem, který se bude hojně vyskytovat v kapitole 3.2 je konvoluce [35], respektive **konvoluční vrstva** [1]. V případě, že neuronová síť obsahuje tuto konvoluční vrstvu, je referována jako konvoluční neuronová síť. Tato vrstva slouží k počítačovému zpracování obrazu, a jejím hlavním přínosem je schopnost detekce prvků na snímku, jako jsou světlé nebo tmavé (nebo specificky zbarvené) body, okraje v různých orientacích, vzory atd. Tyto identifikované prvky následně odhalují významné objekty na obrázku, například lidské oko, nos, ústa a další.

Hlavním cílem konvoluční vrstvy je tedy schopnost rozeznat objekty na obrázcích, nezávisle na jejich umístění, natočení a osvětlení. To zároveň snižuje velikost potřebné trénovací sady. Konvoluční vrstvy se často nacházejí až ve spodních vrstvách neuronových sítí [1].

Podle vzoru lidského mozku se neuronová síť skládá z tisíců nebo dokonce milionů jednoduchých výpočetních uzlů – neuronů, které jsou hustě propojeny. Většina dnešních neuronových sítí je uspořádána do vrstev uzlů a pracuje podle principu *feed-forward*. Na obrázku 3.2 lze vidět, že jednotlivý uzel může být propojen s několika uzly ve vrstvě nižší (podle obrázku se jedná o uzly nalevo od daného uzlu), od kterých přijímá data, a s několika uzly ve vrstvě vyšší (uzly napravo od daného uzlu), kam data posílá [12].

3.2 Generativní neuronové sítě

Generativní neuronové sítě [13] patří mezi klíčové nástroje v oblasti strojového učení. Jedná se o neřízené nebo částečně řízené strojové učení pro vytváření nového obsahu – především v oblasti digitálních obrázků, videí, zvuku, textu a zdrojového kódu. Tyto sítě jsou navrženy tak, aby mohly vytvářet nový obsah na základě vzorů a dat, která byla před nimi prezentována. Jsou to modely, které se snaží pochopit a replikovat vzory v datech, aby mohly generovat nový obsah, který je podobný témtoto vzorům, ale zároveň je i originální.

Jedním ze zástupců generativní umělé inteligence dosáhl v posledních letech vrcholu popularity, mluvíme o *Generative Pre-trained Transformer*, více známý pod zkratkou **GPT**, u kterého se jedná o auto-regresivní jazykový model založený na architektuře transformátoru. Je dopředu trénovaný, ale není již pod dohledem (není supervizován). Tento model je známý především díky produktu GPT-3.5 (a GPT-4) od společnosti OpenAI, spuštěný v roce 2022 [13].

Mimo GPT zde existuje další velký sektor umělých neuronových sítí – jedná se o *Generative Adversarial Network* (Generativní adverzní sítě) ve zkratce **GAN** [9]. Tyto sítě využívají dvě neuronové sítě, které umožňují si navzájem předávat informace pro učení a tím zdokonalovat model. Princip konkurenčního učení mezi těmito sítěmi vedl k významným pokrokům v oblasti generativního umění, což zahrnuje generování obrázků, videa a audia s vysokou úrovní autentičnosti a estetiky. GAN sítě tak představují další mocný nástroj pro tvorbu obsahu, který umožňuje vytvářet realistické a nápadité digitální výstupy [9][13].

Jejich použití je možné vidět v oblasti marketingu a reklamy, kde generativní modely mohou pomoci vytvářet personalizovaný obsah pro zákazníky, ve video produkci pro tvorbu realisticky vypadajících videí přizpůsobených potřebám tvůrce, v hudbě pro generování nových hudebních kompozic a zvukových efektů nebo v umění a kultuře pro tvorbu originálních uměleckých děl.

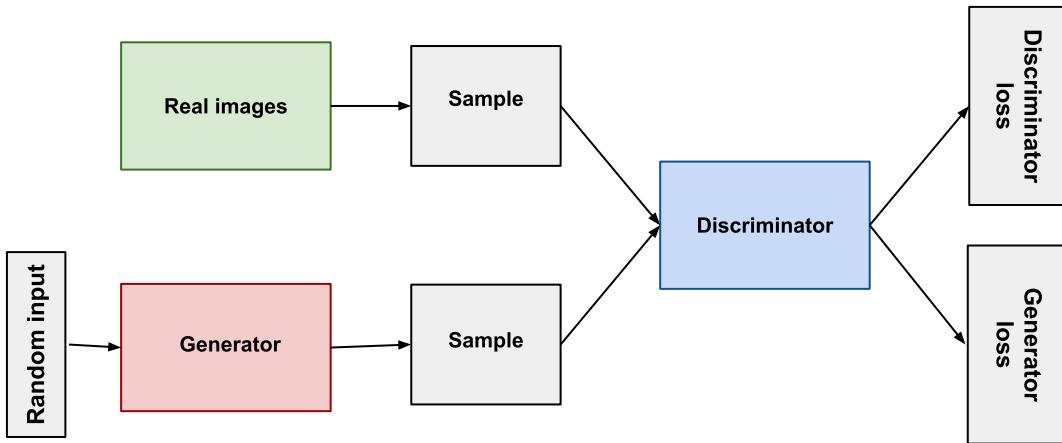
Generativní adverzní sítě

Generative Adversarial Network [9] je rámec strojového učení pro tvorbu vizuálního obsahu. Využívá dvě neuronové sítě, které spolu „soutěží“ a vzájemně se zlepšují prostřednictvím svého soupeřivého vztahu. První neuronovou sítí je **generátor** (G), který zachycuje distribuci dat, a k němu **diskriminační model** (D), který odhaduje pravděpodobnost, zda vzorek pochází z trénovacích dat, a ne z modelu G .

Oba modely se navzájem zlepšují. Model G poskytuje modelu D data, která D využívá k přesnějšímu určení věrohodných informací, respektive model D se učí rozeznat, jestli vzorek

pochází z distribučního modelu G , nebo z předem určené datové sady (vizuálně jde proces vidět na obrázku 3.3). Hlavním cílem modelu G je však maximalizovat pravděpodobnost, že D udělá chybu. Tímto způsobem se G snaží stát více věrohodným modelem. Model D poskytuje zpětnou vazbu ohledně věrohodnosti generovaných dat a tím posiluje věrohodnost modelu G , což vede k vytváření kvalitnějších a spolehlivějších dat.

Lze také použít analogii, ve které je generativní model G skupina padělatelů, kteří se snaží vytvořit falešné peníze, aniž by se na to příšlo. Diskriminační model D reprezentuje policejní trestní orgán, která se snaží odhalit tyto padělané peníze. Obě strany mají motivaci zlepšovat se, s cílem předčít své vzájemné protějšky [9][10].



Obrázek 3.3: GAN diagram obsahující model generátoru, generující výsledky z dat získaných na vstupu a diskriminační model, určující věrohodnost získaných vstupů od generátoru a datové sady. Postavené na bázi hry pro 2 – minimax [10].

Celý proces odpovídá algoritmu minimax mezi 2 hráči. V případě, že jsou G a D definovány pomocí vícevrstvých perceptronů, lze celý systém trénovat pomocí *backpropagation* algoritmu (zpětné šíření), jejímž prostřednictvím poskytuje diskriminační klasifikátor D signály, které generátor G používá k aktualizaci svých vah [9][10].

Zmíněná minimax hra pro dva hráče (D a G) s hodnotovou funkcí $V(G, D)$, lze vyjádřit vzorcem

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (3.1)$$

kde G je generativní model a D je model diskriminační, v obou případech je model vícevrstvý perceptron, v opačné situaci není možné rámec použít. x představuje vzorky reálných dat a z představuje vstupní proměnné šumu, které jsou vzorkovány z rozdělení s prioritou. $p_{\text{data}}(\mathbf{x})$ označuje skutečné rozdělení dat, $p_{\mathbf{z}}(\mathbf{z})$ označuje rozdělení šumu. Výstupem rovnice (3.1) je jeden skalární výraz.

Loss function se vyskytuje jak u generativního modelu G , tak u modelu diskriminačního D . U generátoru označuje, do jaké míry se podařilo oklamat diskriminační model pomocí vygenerovaného výsledku. Na straně diskriminačního modelu jde o to, jak efektivně dokázal odhalit rozdíl mezi reálným vzorkem a tím, který byl vytvořen generátorem. Další obecné informace k pojmu *loss function* se nachází v sekci 3.1.

GAN je rámcem generativních neuronových sítí, umožňující generování vizuálního a audio obsahu. Je složen ze dvou neuronových sítí, které proti sobě soupeří ve hře minimax a navzájem se zdokonalují. Výhody zahrnují to, že nikdy není potřeba Markovův řetězec, během učení není potřeba žádná odvozovací metoda, k získání gradientů se používá pouze zpětné řízení a do modelu lze začlenit širokou škálu funkcí. Nevýhodou je, že D musí být během trénování dobře synchronizované s G , zejména G nemůžeme být příliš trénováno, aniž by se neaktualizovalo D . To může způsobit, že trénink bude nestabilní a pomalý [9].

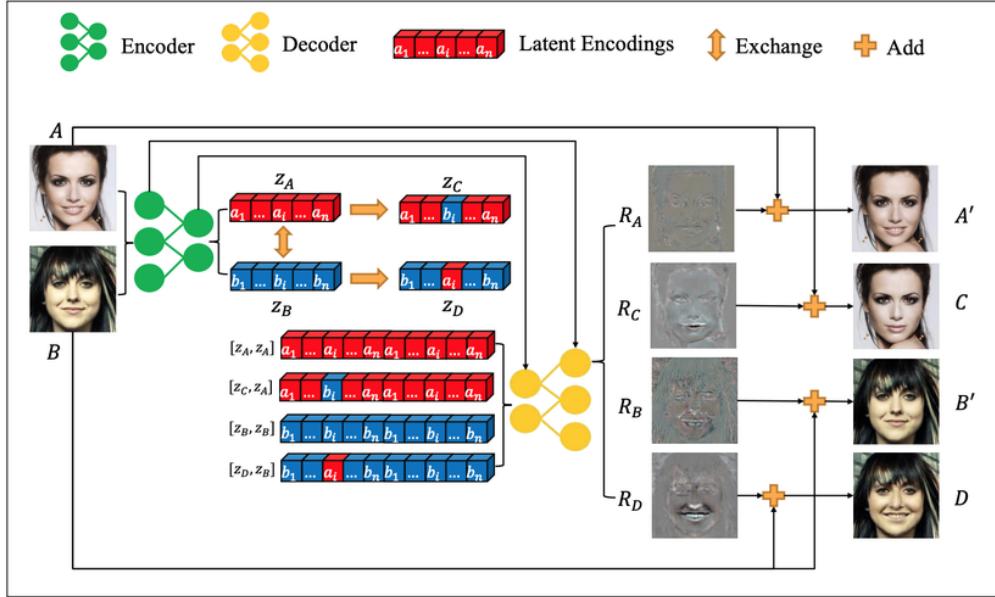
3.3 Generování syntetických snímků obličeje umělou inteligencí

V současné době se GAN používají v různých oblastech výzkumu a aplikací, jako je generování obrazu, překreslování obrazu, generování textu, zpracování lékařských snímků, segmentace obrazu, barevné zobrazení, převod obrazu na obraz a tvorba uměleckých děl. Kromě toho se GAN hojně využívají při **syntéze a úpravě obličejů**, například při určování věku a pohlaví osob.

V posledních letech je syntéza snímků obličeje žhavým tématem v oblasti zpracování fotografií, vzhledem k častému využívání snímků na sociálních sítích. Zpracování obrazů obličeje dosáhlo díky zlepšení výkonu GAN velkého pokroku. Objevila se řada metod pro zlepšení kvality generování obrazu obličeje. Následující modely, představují současný stav generování lidského obličeje umělými neuronovými sítěmi. Často tyto modely pracují na základě zpracování atributů v obrázku a jejich úpravě [34].

ELEGANT

Nástroj ELEGANT poskytuje model, který umožňuje přenos více atributů obličeje, umožňující na snímku lidské tváře například odstranění ofiny, změny barvy vlasů nebo změny mužských rysů na ženské. Schéma rámce lze vidět na obrázku 3.4.



Obrázek 3.4: Struktura modelu ELEGANT pro změnu atributů a následnou syntézu snímků obličeje [7].

Loss function diskriminátoru D

$$L_D = L_{D_1} + L_{D_2} \quad (3.2)$$

kde L_D značí ztrátu diskriminátoru, L_{D_1} a L_{D_2} značí ztráty více-škálových diskriminátorů [7].

Loss function generátoru G

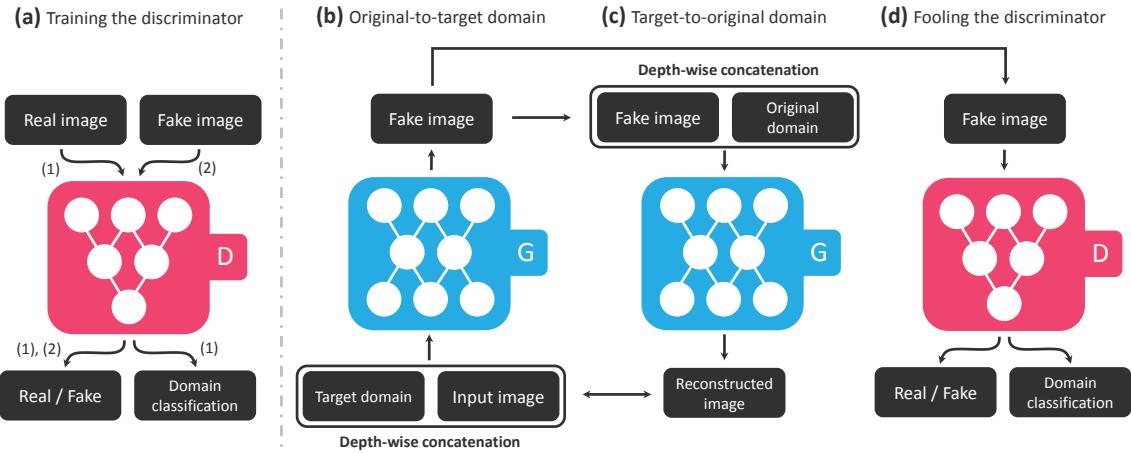
$$L_G = L_{reconstruction} + L_{adv} \quad (3.3)$$

kde L_G značí ztrátu generátoru, $L_{reconstruction}$ ztrátu rekonstrukce a L_{adv} nepříznivou ztrátu [7].

ELEGANT představuje účinnou metodu pro přenos atributů obličeje. Jako vstupy přijímá dva obrazy s opačnými atributy a dokáže vytvořit vysoce kvalitní snímky s drobnými detaily. Dále umožňuje výměnu specifických kódových částí, což umožňuje přenést stejné atributy z jednoho obrázku na druhý. Tato metoda může manipulovat s několika atributy současně tím, že zakóduje různé atributy do nespojitých částí v latentním prostoru. To znamená, že je možné provést změny například v barvě vlasů a současně upravit tvar nosu [34].

StarGAN

StarGAN [2] je model vytvořený Y. Choi, umožňující vytvářet nové snímky v rámci multidoménového překladu, za pomocí jen jednoho modelu. Tato vlastnost poskytuje ještě flexibilnější možnost generování obrázků.



Obrázek 3.5: Přehled sítě StarGAN [2].

StarGAN se skládá ze dvou částí – generátoru G a diskriminátora D . D je trénován nad sadou obrázků, štítkovaných s informací o jejich původu. G se snaží generovat obrázky, tak aby D neodhalil, že se nejedná o reálné obrázky.

Loss function modelu StarGAN

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^r, \quad (3.4)$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec}, \quad (3.5)$$

kde \mathcal{L}_D značí ztrátu diskriminátora a \mathcal{L}_G ztrátu generátoru. \mathcal{L}_{adv} představuje nepříznivou ztrátu (*Adversarial Loss*), \mathcal{L}_{cls}^r ztrátu klasifikace domény (*Domain Classification Loss*) a \mathcal{L}_{rec} ztrátu při rekonstrukci (*Reconstruction Loss*). λ_{cls} a λ_{rec} jsou hyperparametry, které řídí relativní důležitost klasifikačních a rekonstrukčních ztrát v doméně v porovnání s nepříznivou ztrátou [2].

Generátor se učí ignorovat nespecifikované označení a zaměřuje se na ty, která byla určena. Architektura tohoto modelu částečně vychází z konceptu modelu CycleGAN. Generátor je utvářen dvěma konvolučními vrstvami s velikostí kroku dva pro *downsampling*, šesti zbytkovými bloky a dvěma transponováními konvolučními vrstvami s velikostí kroku dva pro *upsampling*. Model slouží pro škálovatelný převod *image-to-image* (obrazu na obraz) mezi více doménami pomocí jediného generátoru a diskriminátoru [2].

STGAN

Nástroj *Spatio-Temporal GAN* slouží pro editaci obličejových atributů, dosahující vysoce kvalitní stupeň výsledků úprav.



Obrázek 3.6: Výsledky STGAN pro úpravu atributů obličeje [23].

Loss function modelu STGAN

$$\min_D \mathcal{L}_D = -\mathcal{L}_{D_{adv}} + \lambda_1 \mathcal{L}_{D_{att}} \quad (3.6)$$

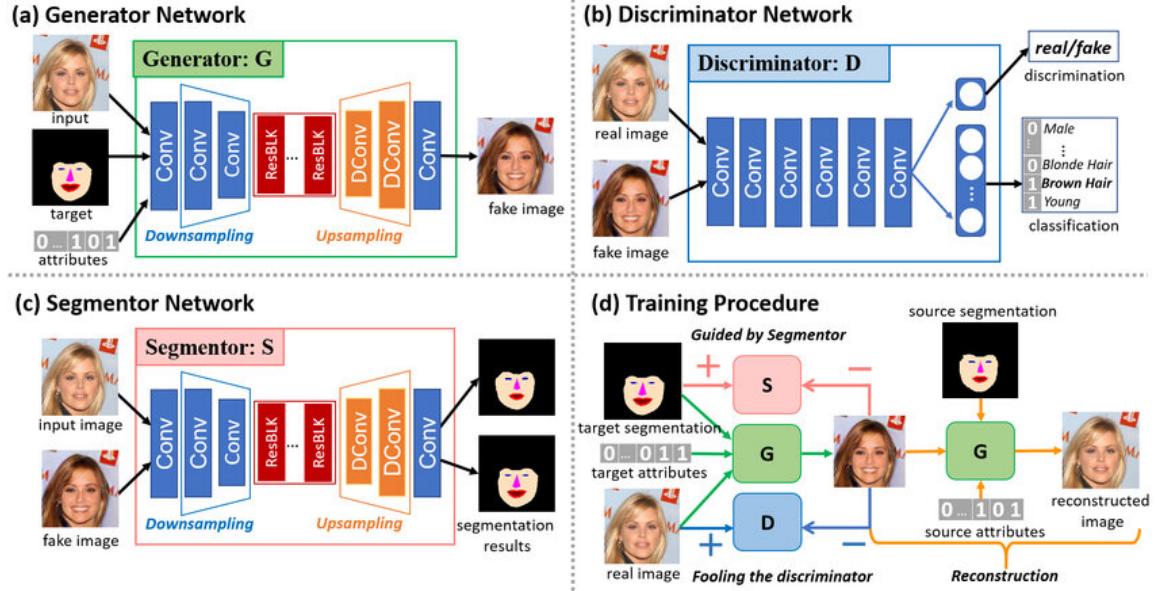
$$\min_G \mathcal{L}_G = -\mathcal{L}_{G_{adv}} + \lambda_2 \mathcal{L}_{G_{att}} + \lambda_3 \mathcal{L}_{rec} \quad (3.7)$$

kde \mathcal{L}_D značí ztrátu diskriminátoru a \mathcal{L}_G ztrátu generátoru. $\mathcal{L}_{D_{adv}}$ a $\mathcal{L}_{G_{adv}}$ jsou nepříznivé ztráty, $\mathcal{L}_{D_{att}}$ a $\mathcal{L}_{G_{att}}$ jsou ztráty při manipulaci s atributy a \mathcal{L}_{rec} je rekonstrukční ztráta. λ_1 , λ_2 , λ_3 jsou parametry kompromisu modelu [23].

Tato metoda řeší jemnou kontrolu označení atributu obličeje a realizuje transformaci více atributů. Model bere jako vstup vektor rozdílových atributů a v konkrétních editačních úlohách mění související atributy namísto všech cílových atributů. STGAN je vysoce přesný model pro editaci atributů založený na modelu StarGAN [2]. Autoři navrhli selektivní přenosové jednotky (*selective transfer units*) pro zlepšení editace atributů, které mohou zlepšit přesnost manipulace s atributy a zlepšit kvalitu zobrazení. STGAN může zlepšit kvalitu generovaných obrazů a realizovat flexibilní překlad atributů tím, že se primárně zaměří na úpravu atributů, které mají být změněny [34].

SGGAN

Model *Segmentation Guided Generative Adversarial Networks* od autora S. Jiang pro překlad obrazu obličeje s více doménami.



Obrázek 3.7: Struktura modelu SGGAN – a) generátor G, b) diskriminátor D, c) segmentor S, během tréninku S poskytuje G prostorové pokyny, aby se zajistilo, že generované obrazy odpovídají vstupním segmentacím, d) tréninková procedura zajišťující realistiky výstup [8].

Loss function modelu SGGAN

$$\mathcal{L}_S = \mathcal{L}_{seg}^{real} \quad (3.8)$$

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{cls}^{real} \quad (3.9)$$

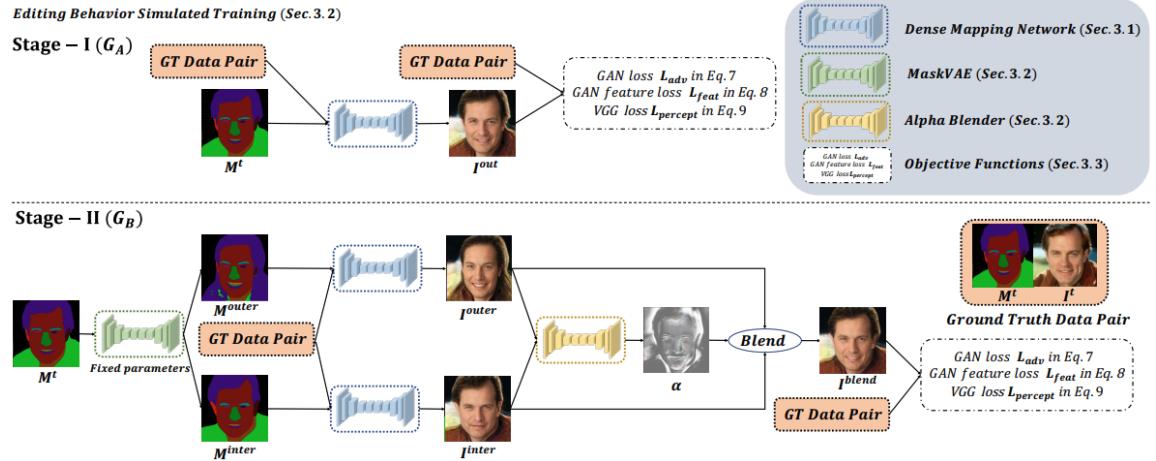
$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{cls}^{fake} + \lambda_2 \mathcal{L}_{seg}^{fake} + \lambda_3 \mathcal{L}_{rec} \quad (3.10)$$

kde \mathcal{L}_S je ztráta segmentoru sítě, \mathcal{L}_D je ztráta diskriminátoru a \mathcal{L}_G je ztráta generátoru. \mathcal{L}_{seg}^{real} a \mathcal{L}_{seg}^{fake} jsou segmentační ztráty, \mathcal{L}_{adv} je nepříznivá ztráta. \mathcal{L}_{cls}^{real} a \mathcal{L}_{cls}^{fake} jsou klasifikační ztráty a \mathcal{L}_{rec} je rekonstrukční ztráta. $\lambda_1, \lambda_2, \lambda_3$ jsou hyperparametry, které řídí váhy klasifikační ztráty, segmentační ztráty a rekonstrukční ztráty [8].

Metoda je založena na hlubokém generativním modelu, který věnuje pozornost informacím vyšší úrovně a specifickým informacím o instancích a dokáže generovat realistiké obrazy vysoké kvality. Má prostorovou kontrolu v procesu překladu obrazu tím, že využívá sémantickou segmentaci ke zlepšení výkonu generování obrazu a poskytuje prostorové mapování. Umožňuje zlepšit kvalitu generování obrazu díky schopnosti prostorové modifikace [34].

MaskGAN

Manipulace s obrazy obličeje dosáhla v posledních letech velkého pokroku. Předchozí metody však budou pracují s předem definovanou sadou atributů obličeje, nebo ponechávají uživatelům jen malou volnost při interaktivní manipulaci s obrázky. K překonání těchto nedostatků navrhujeme nový rámec označovaný jako MaskGAN, který umožňuje rozmanitou a interaktivní manipulaci s obličeji [22].



Obrázek 3.8: Celkový tréninkový proces u MaskGAN. Úprava chování simulovaného tréninku lze rozdělit do dvou fází. Po načtení předtrénovaného modelu husté mapovací síť a MaskVAE tyto dvě fáze iterativně aktualizujeme, dokud model nekonverguje [22].

Loss function modelu MaskGAN:

$$\mathcal{L}_{G_A, G_B} = \mathcal{L}_{adv}(G, D_{1,2}) + \lambda_{feat} \mathcal{L}_{feat}(G, D_{1,2}) \quad (3.11)$$

$$+ \lambda_{percept} \mathcal{L}_{percept}(G) \quad (3.12)$$

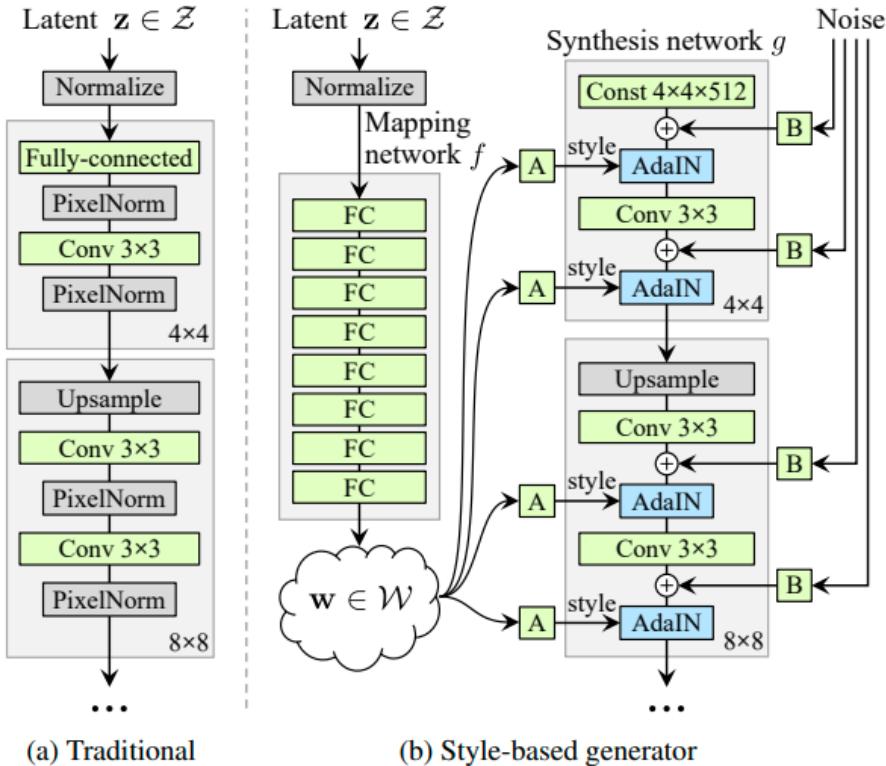
kde G je generátor, D je diskriminátor a \mathcal{L}_G je ztráta generátoru. \mathcal{L}_{adv} je podmíněná nepříznivá ztráta, která zvyšuje realističnost generovaných obrazů a opravuje strukturu generování. \mathcal{L}_{feat} je ztráta při porovnávání prvků a $\mathcal{L}_{percept}$ je percepční ztráta, která zlepšuje generování obsahu od nízkofrekvenčních k vysokofrekvenčním detailům ve vnímání směrem k hlubokým rysům. λ_{feat} a $\lambda_{percept}$ jsou parametry pro úpravu ztráty [22].

Tato metoda překonává omezení spojená s práci s předem definovaným souborem obličejo-vých atributů. Využití sémantických masek jako reprezentace umožňuje uživatelům manipulaci s obrazovými daty obličejů s větší svobodou a interaktivitou. Rámec MaskGAN umožňuje flexibilní manipulaci s obličejo-vými snímkami a zároveň zachovává věrnost původnímu obrazu [34].

V rámci vývoje MaskGANu byl vytvořen rozsáhlý soubor dat o tvářích s vysokým rozlišením a jemnými anotacemi masek, nazvaný CelebAMask-HQ [22], který vychází z původní datové sady CelebA [24].

StyleGAN

StyleGAN [17] je model pro generování obrázků, úpravou podoby originálního obrázku a to bez potřeby supervize. Model pracuje s klíčovými parametry obličeje, jako je výraz tváře, délka a barva vlasů, barevný odstín pokožky, věk, pohlaví, a přítomnost vnějších objektů, například brýlí nebo šperků. Bere v potaz i natočení hlavy. Autorem tohoto modelu je T. Karras. Předtrénovaný model společně s doprovodným datasetem je volně dostupný¹.



Obrázek 3.9: Na obrázku je možné vidět porovnání tradičního generátoru (a) a stylově založeného generátoru (b), kterým je StyleGAN [17].

V procesu modelu StyleGAN se prvotně vstup mapuje do meziprostoru \mathcal{W} , který následně řídí generátor prostřednictvím AdaIN (*adaptive instance normalization*) v každé konvoluční vrstvě. V dalším kroku je přidán Gaussův šum (*noise*). Zde „A“ znamená naučenou affinní transformaci a „B“ aplikuje na vstupní šum naučené škálovací faktory pro každý vstupní kanál. Mapovací síť f se skládá z 8 vrstev. Syntetická síť g se skládá z vrstev 18 — dvě pro každé rozlišení ($4^2 - 1024^2$). Výstup poslední vrstvy je převeden na RGB pomocí samostatné 1×1 konvoluce [17].

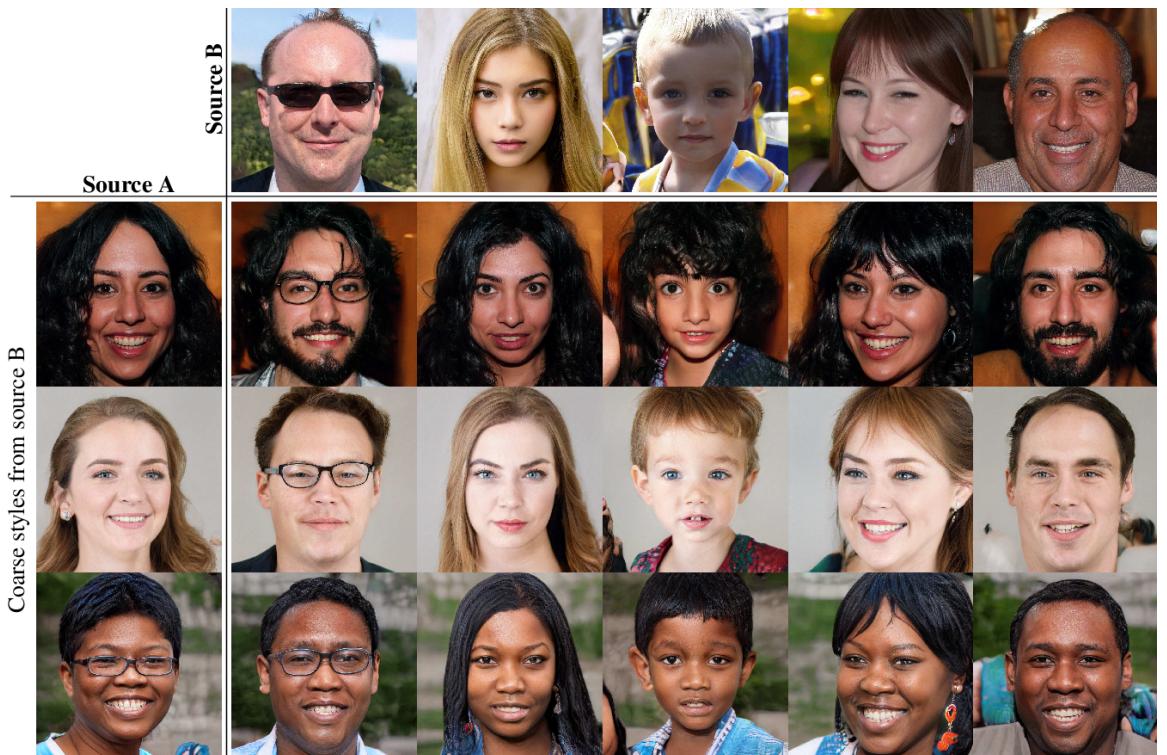
Tento model se odkláňá od tradičního pojetí, ve kterém je latentní kód generátoru poskytován prostřednictvím vstupní vrstvy. StyleGAN využívá mapování latentního kódu z pomocí mapovací sítě f do latentního prostoru \mathcal{W} , kde jsou dále data dostupná pro každou konvoluční vrstvu. Rozdíl jde vidět na obrázku 3.9. Naučená affinní transformace poskytnutá od latentního prostoru určuje styl pro *adaptive instance normalization* (AdaIN), tato operace je definování jako

¹<https://github.com/NVlabs/stylegan>

$$AdaIN(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i} \quad (3.13)$$

kde \mathbf{x}_i je vstupní obrazový vektor a \mathbf{y} je vektor stylu. $\mathbf{y}_{s,i}$ je skalární faktor stylu a $\mathbf{y}_{b,i}$ je vektor posunu, oba pro použití v adaptivní normalizaci s \mathbf{x}_i . $\mu(\mathbf{x}_i)$ značí průměr \mathbf{x}_i a $\sigma(\mathbf{x}_i)$ směrodatnou odchylku \mathbf{x}_i . Každé \mathbf{x}_i je normalizováno samostatně a poté je škálováno na základě stylu \mathbf{y} . V poslední řadě je generátoru poskytnut vstup šumů, pro vytvoření stochastických detailů [17].

Tento model se zaměřuje na možnost modifikování stylu specifickým škálováním, která umožňuje různé zaměření na vykreslení detailů. Další vlastností je míchání stylů (*style mixing*), které umožňuje spojit dva různé styly, místo jednoho. Děje se tak při změně jednoho latentního kódu na druhý v náhodně vybraném bodě sítě syntézy. Přesněji existují dva latentní kódy $\mathbf{z}_1, \mathbf{z}_2$, které odpovídají latentním prostorům $\mathbf{w}_1, \mathbf{w}_2$. Styl \mathbf{w}_1 je aplikován až do okamžiku výměny, kde pak přebírá kontrolu styl \mathbf{w}_2 [17].



Obrázek 3.10: Dvě sady snímků byly vygenerovány z odpovídajících latentních kódů (*Source A* a *Source B*). Zbytek obrázků byl vygenerován pomocí stylu z zdroje B a následně zdroje A. Kopírování stylů odpovídající hrubému rozlišení ($4^2 - 8^2$), umožňuje přenést vlastnosti vyšší perspektivy, jako jsou póza, styl účesu, tvar obličeje a brýle, ze zdroje B a styl barev (oči, vlasy, nasvícení) je přenesen ze zdroje A [17].

StyleGAN2 [18] představuje novou generaci tohoto modelu, vyvinutou s cílem eliminovat problematické oblasti, které se objevily v původní verzi StyleGAN [17]. Hlavním problémem, který byl řešen, je odstranění rušivých prvků, aby bylo dosaženo generování snímků, které jsou ještě realističtější a obtížně odlišitelné od skutečných fotografií. Jedním

z těchto rušivých prvků byly artefakty připomínající vodní kapky, které vznikaly díky normalizaci instancí. StyleGAN2 proto omezuje použití adaptivní normalizace instancí s cílem odstranit tyto rušivé prvky.

StyleGAN3 [16] představuje pokračování této výzkumné práce. Na rozdíl od StyleGAN2 má několik vylepšení v síťové struktuře a v metodách tréninku. Hlavním zdokonalením je snaha zajistit, aby animace přechodů byla přirozenější, a konkrétně řeší problémy s ulpíváním textur při generování. Jedním z hlavních nedostatků vytvořených sérií snímků bylo nedostatečně realistické zobrazení vlasů a vousů.

Generativní adverzní síť tří hráčů

GAN síť tří hráčů, přináší do minimax hry mimo generátor a diskriminátor ještě třetího člena, který se zaměřuje na identitu (ID-3 [20]), toto rozšíření poskytuje model IDnet. Hlavním cílem modelu IDnet je problematika hry dvou hráčů při generování syntetických snímků obličeje bez oddělení identity reálných osob, která může narušovat jejich soukromí a porušovat legislativu, například při tvorbě datasetů pro trénování modelů pro rozpoznání obličeje (*face recognition*). Rozšíření o třetího hráče může zajistit, že k takovým narušením nedojde a vygenerované výsledky budou oddělené od reálných identit lidí, kteří nemuseli poskytnout pro zpracování do datasetu souhlas a to se zachováním kvality výsledných obrázků.

Model IDnet funguje na architektuře dvou hráčů StyleGAN2-ADA s přidáním třetího hráče ID-3, který funguje jako model pro *face recognition*, který je zaintegrován do *loss function*. ID-3 je trénován pomocí *softmax loss function* s penalizací za marži, která zahrnuje penalizaci za marži v softmax ztrátě, aby se trénovací vzorky přiblížily ke středům svých tříd a vzdálily se od středů ostatních tříd. Konkrétně se k tréninku ID-3 používá *loss function face recognition* modelu CosFace. Nás nový hráč má za cíl vést generátor G k tomu, aby se naučil vytvářet syntetické obrazy s vysokou variabilitou a silnou oddělitelností identit [20].

Zhodnocení současného stavu GAN

Neuronové sítě a svět umělé inteligence jsou v současné době na vrcholu, kde rozvoj pokroku a inovací směřuje rychlým tempem kupředu a otevírají se nové možnosti. Konkrétně bylo nahlízeno na oblast generativních neuronových sítí, zejména síť GAN. Tyto sítě pracují s principem soutěže mezi generátorem a diskriminátorem, což vede k vytváření realistických a kvalitních vizuálních obsahů. Existuje mnoho rozdílných přístupů využití GAN pro různé účely, přičemž se nejvíce liší ve ztrátových funkcích. Každý z přístupů má trochu jiné vlastnosti a zpracování snímků pro editaci a tvorbu, nicméně všechny přináší kvalitu a posun ve tvorbě uměle vygenerovaných snímků obličeje. V rámci této práce se zdá nejvhodnější použít model StyleGAN, umožňující generování snímků na škále rozlišení, která přenáší na nově vygenerované snímky různé vlastnosti v závislosti na zvoleném rozlišení.

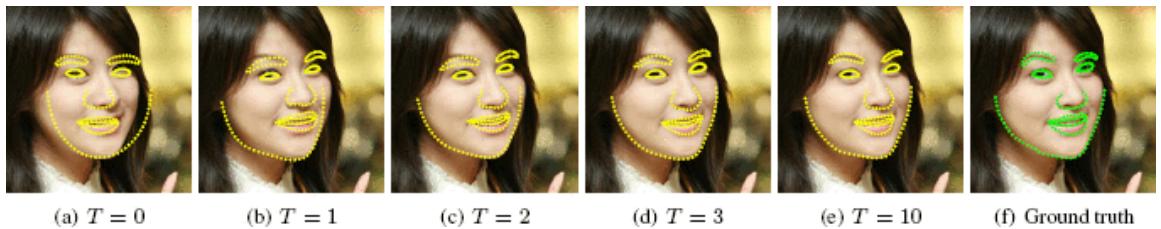
3.4 Detekce antropometrických bodů v obrazu

Důležitým aspektem pro zkoumaní a měření obličeje jedince, v oblasti počítačového vidění, jsou antropometrické body na obličeji, které umožňují význačné oblasti obličeje podrobně zkoumat. Před jakýmkoliv zkoumáním antropometrických bodů je nejprve potřeba lokalizovat obličeje, poté je možné se posunout k detekci těchto bodů. K tomu může být použita například knihovna *OpenCV* pro funkce v oblasti počítačového vidění v reálném čase [29].

Existuje celá řada detektorů obličejoých antropometrických bodů, ale všechny metody se v podstatě snaží lokalizovat a označit následující oblasti obličeje:

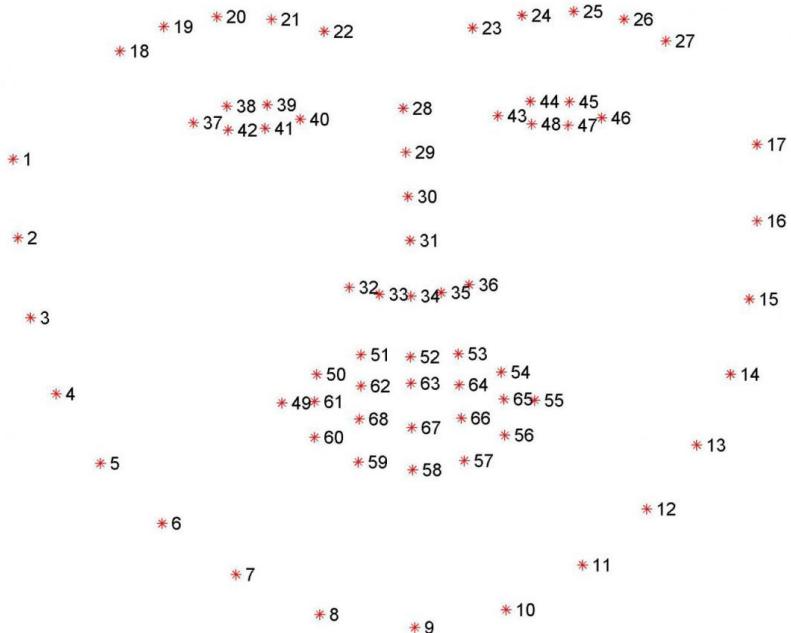
- ústa,
- pravé obočí,
- levé obočí,
- pravé oko,
- levé oko,
- nos,
- čelist.

Knihovna dlib [19][29] je jedním z možných řešení pro vyhledání antropometrických bodů na obličeji, využívá práci autora V. Kazemi a J. Sullivan (dostupná na [19]). Tento rámec je založen na učení se sady regresních stromů pro optimalizaci součtu kvadratických ztrát a přirozeně se vypořádává s chybějícím nebo částečně označenými daty. Algoritmus manuálně označí body na obličeji, podle obecného schématu a poté zdokonaluje jejich polohu x a y souřadnic, dokud antropometrické body nejsou umístěny na správném místě. Proces je zachycen na snímku 3.11.



Obrázek 3.11: Regresivní kaskáda pro určování odhadovaného rozložení antropometrických bodů na obličeji v jednotlivých stádiích. Již po první úrovni kaskády se chyba výrazně sníží [19].

S použitím datové sady (angl. *dataset*) iBUG 300-W, na kterém byl dlib trénován získáme 68 souřadnic (x, y), které mapují obličejové struktury na obličeji. Dataset vznikl v rámci studie: *300 Faces in-the-Wild Challenge* (výzva 300 tváří v divočině), obsahuje autentické snímky z běžného života zachycené na různá zařízení. Dataset byl vytvořen seskládaní fotografií z jiných datových sad (jako HELEN, AFW, LFW a další), včetně úplně nových snímků pro obohacení sady [30]. Základní anotaci datového setu lze vidět na obrázku 3.12.



Obrázek 3.12: 68 (x,y) souřadnic pro poskytnutí anotace datovou sadou iBUG 300-W [30].

MediaPipe

MediaPipe [26] je rámec pro vytváření rour (*pipeline*) pro provádění odvozování nad libovolnými smyslovými daty. Pomocí MediaPipe lze sestavit percepční *pipeline* ve formě grafu modulárních komponent. Do grafu vstupují smyslová data, jako jsou zvukové a video toky, a generuje výstupní percepční popisy, jako jsou například lokalizace objektů nebo bodů obličeje. Hlavním využitím MediaPipe je rychlé vytváření prototypů percepčních *pipeline* s odvozovacími modely a dalšími opakovaně použitelnými komponentami. Díky jednoduché opakovatelnosti lze komponenty MediaPipe snadno využívat v různých *pipeline* napříč po sobě jdoucími aplikacemi, protože tyto komponenty sdílejí společné rozhraní orientované na data. Každá *pipeline* pak může běžet se stejným chováním na různých platformách, což umožňuje odborníkovi vyvinout aplikaci na pracovních zařízeních a následně ji nasadit například na mobilní zařízení.

MediaPipe se skládá ze tří hlavních částí: rámce pro odvozování ze senzorických dat, sady nástrojů pro hodnocení výkonnosti a sady opakovaně použitelných komponent pro odvozování a zpracování dat, tzv. kalkulátorů. *Pipeline* je definována jako směrovaný graf komponent, kde každá komponenta je kalkulátor **Calculator**. V grafu jsou kalkulátory propojeny datovými toky **Streams**. Každý tok představuje časovou řadu datových paketů **Packets**. Kalkulátory a toky společně definují graf datových toků. Pakety, které protékají grafem, jsou v rámci časové řady srovnány podle svých časových značek. *Pipeline* lze postupně vylepšovat vkládáním nebo nahrazováním kalkulátorů kdekoli v grafu. Vývojáři mohou také definovat vlastní kalkulátory [26].

Příkladem použití rámce MediaPipe je detekce objektů v reálném čase. Je například možné zpracovat výstup z videokamery a identifikovat objekty v obraze. Sledovací mechanismus (*tracker*) lokalizuje polohu objektu, zatímco detektor určuje typ objektu. Pro dosažení optimálního výkonu by sledování a detekce měly probíhat paralelně, aby *tracker* nebyl blokován detektorem a mohl zpracovat každý snímek.

Dalším příkladem je detekce antropometrických bodů obličeje. Proces je zde podobný jako u předchozí aplikace, provádí se detekce antropometrických bodů obličeje spolu se segmentací portrétu. Jednou ze strategií, jak snížit výpočetní zátěž potřebnou ke spuštění obou úloh současně, je použít úlohy na dvě oddělené podmnožiny snímků. Výsledná vizualizace je zobrazena jako překrytí na snímcích z kamery [26].

Detekce antropometrických bodů v obraze je zásadním prvkem pro mnoho aplikací v oblasti počítačového vidění, zejména v analýze obličeje jedince. Zatímco knihovna dlib nabízí spolehlivé metody pro lokalizaci těchto bodů, rámec MediaPipe přináší širší perspektivu. Jeho modularita a schopnost práce s různými typy senzorických dat umožňuje rychlé prototypování a nasazení aplikací i v reálném čase. Tím se stává klíčovým nástrojem pro práci s antropometrickými body nejen v analýze obličeje, ale i v širším kontextu počítačového vidění.

3.5 Rozpoznávání a ověřování obličeje

V dnešní digitální době, kdy se technologie neustále rozvíjí, získává rozpoznávání a ověřování obličeje stále větší pozornost. Tento proces, který využívá analýzy obrazu obličeje k identifikaci jednotlivců, se stal klíčovým nástrojem v oblastech bezpečnosti, osobního identifikačního systému a uživatelské autentizace.

Ověřování obličeje je úloha, při níž se na základě analýzy obrazu obličeje rozhoduje, zda je daná osoba tím, za koho se vydává. Tento úkol je velmi náročný vzhledem k rozdílům v osvětlení, póze, výrazu obličeje a věku. Hlavním cílem je vypočítat **vzdálenost** [27] (*distance*) mezi dvěma obličejovými vektory, což představuje míru jejich rozdílnosti.

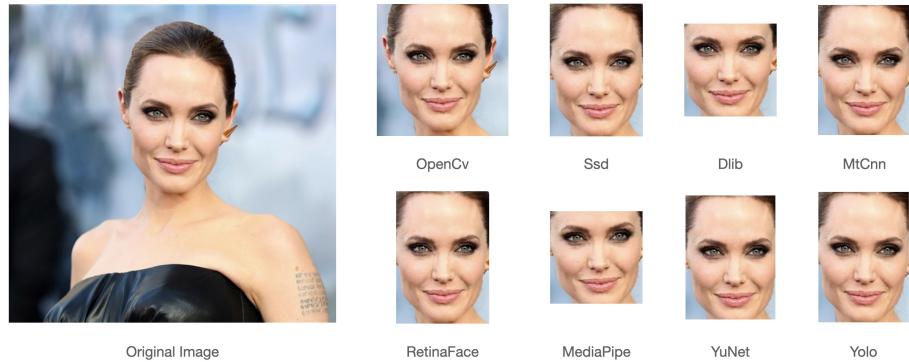
Existuje několik metrik pro měření podobnosti, mezi nimiž se řadí eukleidovská vzdálenost, která měří vzdálenost mezi dvěma vektory, a kosinová vzdálenost, která analyzuje kosinus úhlu mezi také dvěma vektory. Oba tyto přístupy operují ve vícerozměrném prostoru [27].

Deepface

Deepface [31] je odlehčený hybridní *framework* pro rozpoznávání obličeje a analýzu atributů obličeje (věk, emoce, pohlaví, rasa) pro jazyk Python. Jeho hybridní přístup umožňuje přepínat mezi nejmodernějšími modely rozpoznávání obličeje, jako jsou VGG-Face, FaceNet, ArcFace, Dlib a další. Moderní postup rozpoznávání obličeje se skládá z pěti základních fází: detekce, zarovnání, normalizace, reprezentace a ověření. Deepface zvládá všechny tyto fáze na pozadí. Obrázek 3.13 představuje konkrétní příklad detekce obličeje v obraze pomocí vybraných modelů v rámci Deepface.

Tento framework značně usnadňuje práci s moderními (*state-of-the-art*) modely pro rozpoznávání obličeje. Uživatelé mohou jednoduše volat potřebné operace pomocí jediné funkce, které předávají specifické argumenty, jako je výběr metriky, modelu pro rozpoznávání, modelu pro detekci a dalšího nastavení. To znamená, že i uživatelé s omezenými znalostmi hlubokého učení mohou snadno využívat pokročilé technologie rozpoznávání obličeje bez nutnosti detailní znalosti fungování jednotlivých modelů a algoritmů. Kód rámce je dostupný pod licencí MIT na platformě GitHub².

²<https://github.com/serengil/deepface>



Obrázek 3.13: Ukázka různých přístupů modelů využitelných v rámci Deepface k detekování lidského obličeje [31].

ArcFace

ArcFace patří mezi modely vyvinuté pro řešení úkolů v oblasti rozpoznávání obličeje. Na rozdíl od jiných modelů nevyužívá přímo *Softmax loss function*, ale její upravenou verzi nazvanou *Additive Angular Margin Loss*. Tato úprava přináší značné výhody, jako je možnost pracovat s nečistými daty a zachovat přitom vysoký výkon. K výpočtu úhlu mezi aktuálním prvkem a cílovým středem používá funkci arkus kosinus (*arccosine*). Poté k cílovému úhlu zavede *additive angular margin* a pomocí kosinové funkce získá cílový logaritmus. Poté všechny logaritmy přeškáluje pevnou normou funkce. Následné kroky jsou potom prakticky totožné s postupem při využití *softmax loss*. ArcFace přímo optimalizuje rozpětí geodetických vzdáleností (*geodesic distance*) díky přesné korespondenci mezi úhlem a obloukem v normalizované hypersféře. Dále je schopen pracovat s nadměrným šumem a dokáže tato tréninková data řádně vyčistit [6].

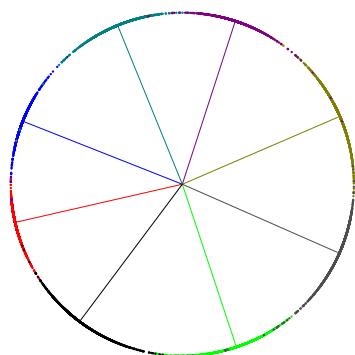
Additive Angular Margin Loss modelu ArcFace

$$L = -\log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + \sum_{j=1, j \neq y_i}^N e^{s \cos \theta_j}}. \quad (3.14)$$

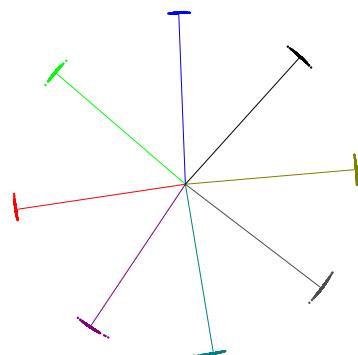
kde x_i je vkládaný rys, W_{y_i} odpovídá třídě y_i , s je měřítko, které reguluje velikost úhlového rozptylu, θ_j je úhel mezi váhou W_j a rysem x_i , N je celkový počet tříd a m je úhlová penalizace mezi W_{y_j} a x_i [6].

ArcFace je inovativním modelem pro rozpoznávání obličeje, který nepožaduje, aby datová sada obsahovala pouze čistá data. Jeho ztrátová funkce se používá k trénování neuronových sítí tak, aby vektory rysů byly lépe rozděleny v prostoru a byly lépe separovatelné podle tříd. Jeho rozšíření *softmax loss function* umožňuje efektivní práci s vyšším šumem a přesto udržuje působivý výkon. Díky této schopnosti představuje ArcFace významný po-

krok v oblasti rozpoznávání obličeje, který nabízí robustní řešení i pro reálné podmínky s nečistými daty.



(a) Softmax



(b) ArcFace

Obrázek 3.14: Příklad rozdílu mezi ztrátovou funkcí (*loss function*) Softmax a ArcFace na 8 různých identitách. Na základě normalizace rysů jsou všechny rysy obličeje posunuty do obloukového prostoru (*the arc space*) s pevným poloměrem [6].

Kapitola 4

Návrh řešení a implementace analýzy antropometrických proporcí

V této kapitole je představen návrh aplikace určené k analýze antropometrických proporcí pomocí pokročilých technologií z oblasti počítačového vidění a umělé inteligence. Hlavním cílem aplikace je posoudit rozdíly mezi uměle generovanými a reálnými snímky obličeje. Řešení aplikace spočívá v kombinaci již existujících metod umělé inteligence a vytvoření integračního systému, který umožní komplexní analýzu dané problematiky. S ohledem na spíše kombinační charakter práce a pouze částečně implementační, se na konci této kapitoly nachází i implementační detaily k jednotlivým použitým procesům.

Technické podrobnosti aplikace:

- Operační systém: macOS 14.2
- Programovací jazyk a knihovny: Python 3.10
 - deepface, MediaPipe, OpenCV, Matplotlib, Numpy

Generování umělých snímků obličeje bude probíhat za pomoci modelu StyleGAN3. Vstupem pro model bude podmnožina datového setu CelebAMask-HQ o velikosti 250 snímků a výstupem budou uměle vygenerované snímky obličeje. Výsledek doplňující dvojice subsetu je základem pro další zpracování a analýzu generativních adverzních modelů. Pro provádění těchto operací jsou nezbytné následující Python knihovny: PyTorch v minimální verzi 1.9.0 a CUDA toolkit ve verzi 11.1 nebo vyšší.

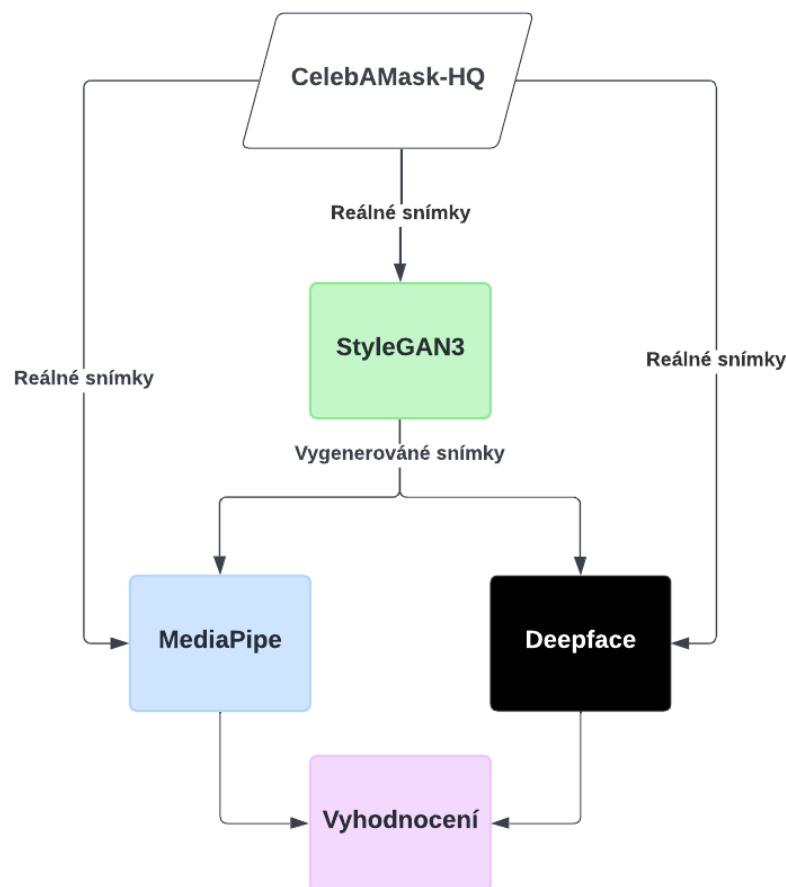
Po získání dvojice dat – reálných a umělých snímků, následuje jejich analýza, která se provádí ve dvou procesech.

Jeden z procesů zajišťuje vyhodnocení podobnosti na základě vzdálenosti vektorů obličeje pomocí modelu DeepFace. Druhý proces – model MediaPipe, detekuje celkovou masku antropometrických bodů na obličeji, která se ve vyhodnocení využije pro výpočet odlišnosti u jednotlivých proporcí antropometrických bodů na obličeji.

4.1 Blokové schéma návrhu

Po seznámení s použitými technologiemi je možné se podívat na jejich propojení v rámci navrženého systému. Diagram blokového schématu je k nalezení na obrázku 4.1.

Data ve formě podmnožiny datové sady CelebAMask-HQ, vstupují do GAN modelu StyleGAN3 pro generování umělých snímků obličeje. Každý vygenerovaný snímek je následně porovnán se svým originálním protějškem pomocí knihovny Deepface. Tento proces je analogický i pro model MediaPipe s důrazem na antropometrickou analýzu obličejů. Výsledky jsou nakonec vyhodnoceny a zobrazeny uživateli.



Obrázek 4.1: Blokové schéma návrhu práce.

4.2 Datová sada reálných snímků obličeje

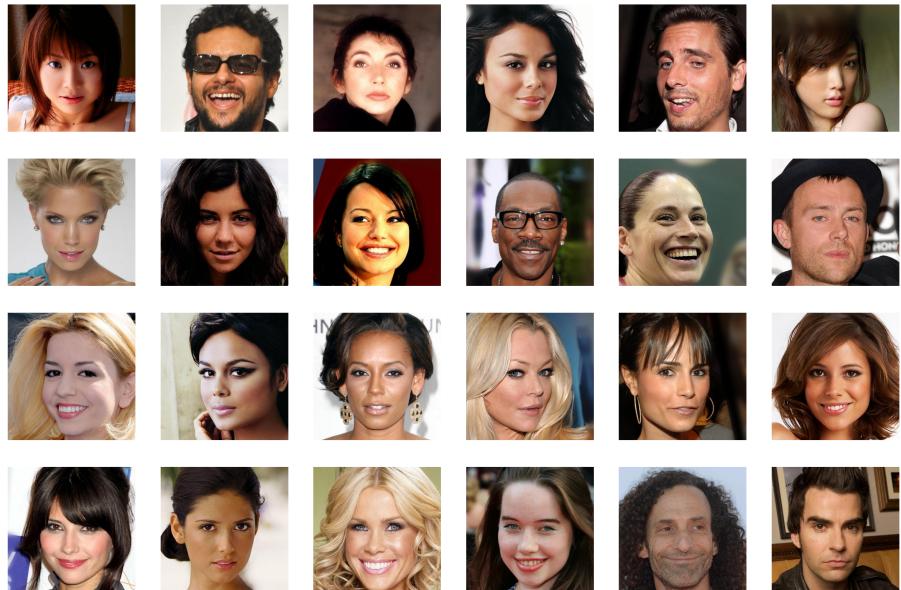
Pro generování nových snímků obličeje a jejich srovnání s původním snímkem je nutné vybrat vhodnou datovou sadu, která disponuje snímkery obličeje skutečných osob. Kromě toho je vhodné z datové sady vytvořit podmnožinu dat, s cílem zlepšit kvalitu vývoje a uskutečňování experimentů.

Oblíbenou volbou je datová sada CelebA Dataset, která zahrnuje 202 599 fotografií obličeju známých osobností a obsahuje více než deset tisíc jedinečných identit. Tato komplexní

datová sada je známa svou pestrostí a obsahuje obličeje s různými charakteristikami, jako jsou brýle, pokrývky hlavy, kudrnaté vlasy, knírky, vousy a variabilitou typů obličejů a dalších atributů. Dataset je příhodný pro širokou škálu aplikací, včetně rozpoznávání atributů obličeje, rozpoznávání obličeje, detekce obličeje a úpravy a syntézy obličeje. Vzhledem k jeho nejnovější aktualizaci v roce 2021 poskytuje aktuální a rozsáhlý soubor dat [24].

Nicméně klasická verze CelebA datasetu může mít nízké rozlišení snímků, což nemusí vždy vyhovovat požadavkům experimentů. Proto bylo vytvořeno rozšíření od autorů původní datové sady s názvem CelebAMask-HQ Dataset, které nabízí snímky s vyšším rozlišením. Obsahuje 30 tisíc snímků vybraných z CelebA datasetu, přičemž každý z nich má rozlišení 1024x1024 pixelů, což přináší výrazné zlepšení kvality obrazu [22]. Jeho podmnožina bude využita v rámci této práce.

Pro výběr podmnožiny datové sady byla využita metoda náhodného výběru s manuální korekcí. Obsah části této podmnožiny je zobrazen na obrázku 4.2. Tento přístup zajišťuje dostatečně velkou a reprezentativní vzorkovou množinu pro účely analýzy a experimentování.



Obrázek 4.2: Část snímků podmnožiny datové sady CelebAMask-HQ.

4.3 Generování umělých snímků obličeje

Existuje mnoho způsobů, které dokážou generovat umělé snímky obličeje, v této práci byl zvolen model StyleGAN3 [16] od výzkumného týmu společnosti NVIDIA. Proces generování je implementován pomocí Python notebooku, který vychází z dostupných příkladů na oficiálním GitHub repozitáři modelu¹. Tento notebook byl dále upraven o možnost vkládání

¹Oficiální repozitář modelu: <https://github.com/NVlabs/stylegan3>

vlastních snímků a následné generování, a to díky rozšířené verzi vyvinuté výzkumníkem Eugeniem Herrera-Bergem, která je dostupná na jeho GitHub repozitáři².

Použitý kód původně pochází od tvůrců a byl rozšířen o další funkce autorem Eugeniem Herrera-Bergou. Pro tuto práci byl notebook rozšířen o další funkce pro lepší použitelnost. Změnily se způsoby zpracování vstupu, kde množství fotografií obličeje není omezeno na jednu, ale uživatel může specifikovat vlastní počet fotografií. Po zpracování každé fotografie je generován umělý snímek pomocí modelu StyleGAN3, a tento proces je prováděn ve smyčce pro všechny zvolené snímky.

Kód v jazyce Python je k dispozici v souboru `StyleGAN3_Generate_images.ipynb`, přičemž samotné generování probíhalo na serverech Google Colab z důvodu náročnosti výpočtu a vyhnutí se problému s použitými knihovnami při spuštění lokálně. Výsledek generování spolu s původním snímkem je zobrazen na obrázku 4.3.

Hlavní viditelný rozdíl ve spektru všech vygenerovaných fotografií spočívá především ve výrazném rozostření pozadí a celkovém odstranění vjemů mimo obličej. Zároveň platí, že čím vyšší je kvalita původních snímků, tím méně se na výsledném obraze objevuje artefaktů a obličeji působí reálněji.



(a) Původní snímek obličeje



(b) Umělý vygenerovaný snímek obličeje

Obrázek 4.3: Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3.

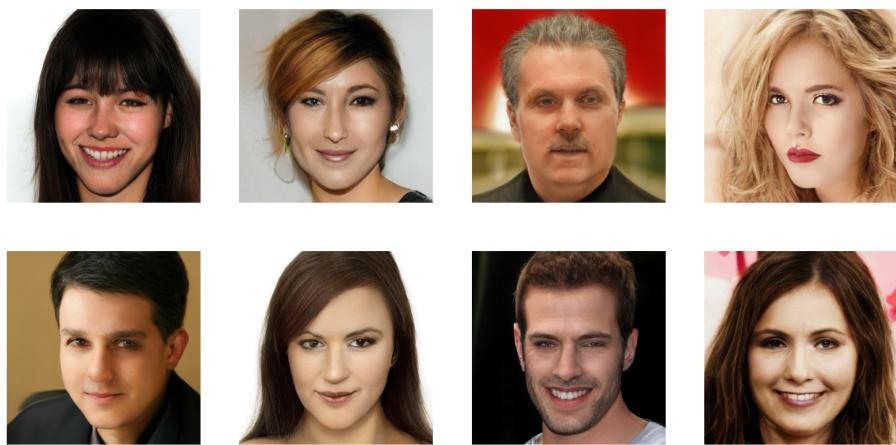
Na začátku v prvním bloku, notebook stáhne všechny nutné závislosti, definuje používané funkce (například pro stažení fotografe) a inicializuje proměnné. V dalším bloku je možné zvolit předtrénovaný model z nabídky modelů. Pro práci je použit model trénovaný na datové sadě Flickr-Faces-HQ [17], což je vysoce kvalitní datová sada snímků lidských tváří, která byla původně vytvořena jako referenční soubor pro generativní adverzní sítě.

V dalším bloku je potřeba provést úpravy v parametrech jako je počet kroků (`steps`), počáteční hodnota `seed` (`seed`) a velikost podmnožiny (`subset_size`), se kterou se bude pracovat. Proměnná `steps` určuje počet kroků, které se provedou při procesu generování obrázků. V kontextu kódu se jedná o počet iterací, které algoritmus provede při projekci vektoru do prostoru obrázků. Čím vyšší hodnota `steps`, tím déle algoritmus trvá, a tím více se model může přizpůsobit cílovému obrázku. Pro generování snímků byla zvolena hodnota

²Rozšířená verze: <https://github.com/ouhenio/stylegan3-projector>

250 kroků. Dále jsou definovány transformace obrazu pomocí knihovny TorchVision (`tf`), která zahrnuje změnu velikosti na 1024x1024 pixelů a normalizaci obrazu do rozsahu [0, 1]. Funkce `run` je definována pro provádění samotného procesu generování obrázků. Nejprve se inicializuje náhodný *seed* pro PyTorch a získají se cílové rysy (`target_features`) pro projekci. Následně se provede samotná projekce, která se skládá z několika iterací. V každé iteraci se vygeneruje několik obrázků, vypočítá se ztrátová funkce (`loss`) a vybere se ten obrázek, který minimalizuje tuto ztrátu. Nakonec se aktualizuje váha (`q`) pomocí optimalizačního algoritmu AdamW. Po dokončení se obrázek zobrazí.

Poslední blok inicializuje cílové obrázky pro projekci. Jsou načteny obrázky ze zadaných URL adres a upraveny do požadované podoby. Poté je spuštěna funkce `run` pro každý cílový obrázek.



Obrázek 4.4: Ukázka vygenerovaných snímků obličeje pomocí modelu StyleGAN3.

Výsledkem této sekce jsou uměle vygenerované snímky, které vytvářejí dvojice s původními reálnými snímky. Tato fáze je klíčová pro následnou analýzu, během níž jsou tyto dvojice porovnávány. Aby se zajistila lepší kvalita výsledného snímku, byly některé z vygenerovaných snímků s viditelně nízkou kvalitou přegenerovány.

4.4 Informace o antropometrických bodech na obličeji

Tento pomocný soubor s názvem `landmark_info.py` obsahuje zásadní informace pro antropometrickou analýzu provedenou pomocí modelu MediaPipe, která je popsána v sekci 4.5. Definuje nezbytné hodnoty, spojení a datové třídy.

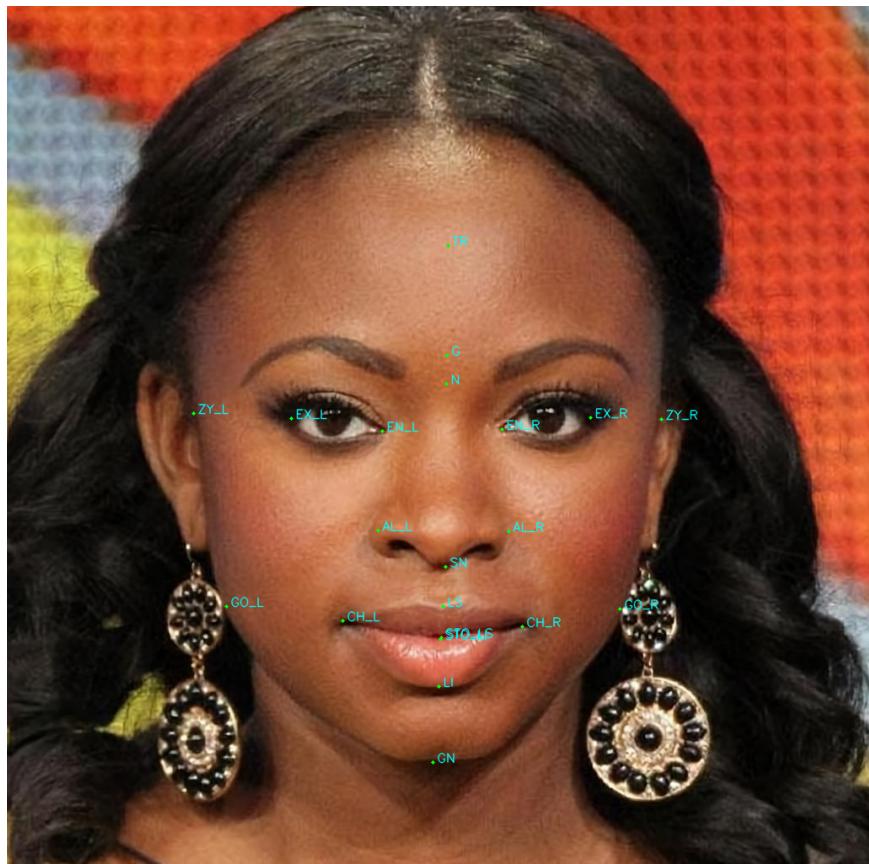
Většina bodů uvedených v tabulce 2.1 je pomocí konstant mapována na indexy jednotlivých bodů na masce obličeje modelu MediaPipe, přičemž bylo vycházeno z popisné fotografie na úložišti Google API pro MediaPipe³, která obsahuje popis všech indexů na vykreslených bodech.

³Dostupné na úložišti Google: https://storage.googleapis.com/mediapipe-assets/documentation/mediapipe_face_landmark_fullsize.png

Tyto body, které je možné vidět na obrázku 4.5, jsou propojovány pomocí datové třídy `LandmarkMeasurements`. Třída obsahuje atributy, kde `start` zastupuje počáteční bod měření, `end` zastupuje koncový bod a `name` představuje název měření. Jednotlivé instance tříd představují konkrétní měření na obličeji a jsou uchovány v konstantě `FACE_PROPORTIONS`, která vytváří seznam těchto vybraných měření. Měření pro analýzu obličeje tvoří podmnožinu měření ze sekce 2.2 a jsou následující:

- měření obličeje – šířka obličeje (zy-zy), šířka čelisti (go-go), výška dolní části obličeje (sn-gn), výška obličeje (tr-gn), výška obličeje (n-gn), výška horní části obličeje (n-sto), výška dolní třetiny obličeje (sto-gn), speciální výška obličeje vlevo (en-gn), speciální výška obličeje vpravo (en-gn), speciální výška horní části obličeje (g-sn)
- měření orbitu – šířka mezi oči vnější (ex-ex), šířka mezi oči vnitřní (en-en), šířka levého oka (ex-en), šířka pravého oka (ex-en)
- měření nosu – šířka nosu (al-al), výška nosu (n-sn)
- měření úst – šířka úst (ch-ch), výška horního rtu (ls-sto), výška dolního rtu (sto-li)

Další implementovanou třídou je datová třída `MaxDifference`, která má atributy `value`, který značí součet všech hodnot jednoho typu měření, a `count`, který představuje počet výskytů tohoto typu. Při každém porovnání dvojic je přidán pouze ten typ měření, který dosáhl nejvyšší hodnoty.

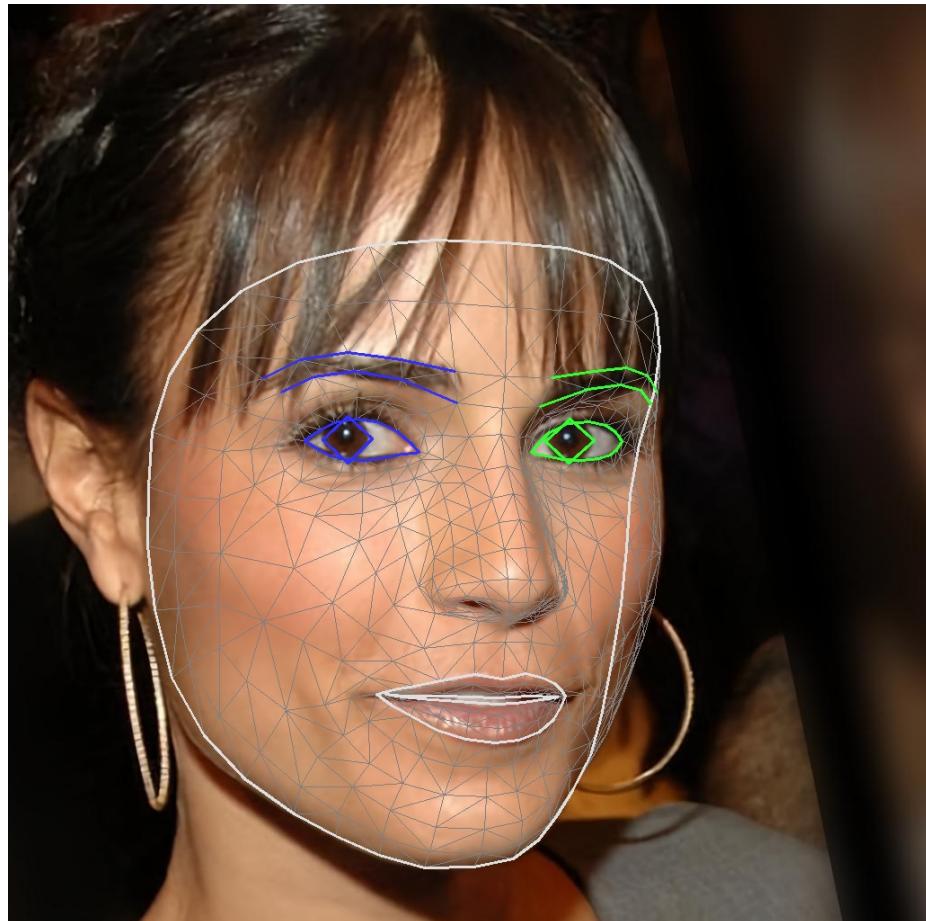


Obrázek 4.5: Vybrané zkoumané antropometrické body na obličeji. Názvy bodů jsou konstanty s hodnotou odpovídající indexu na masce modelu MediaPipe.

Druhý seznam obsahující hodnoty morfologické šířky a výšky obličeje je uložen v konstantě `TFI` pro výpočet poměru `TFI`. Pracuje stále s datovou třídou `LandmarkMeasurements`, ale na rozdíl od předchozí konstanty je její velikost omezena na dvě zmíněná spojení.

4.5 Antropometrická analýza významných bodů obličeje

V této první části analýzy se program zaměřuje na podrobné měření poměrů mezi antropometrickými body na obličeji. Poměry a významné body vychází z antropometrické analýzy provedené v Řecku [36], více informací se nachází v sekci 2.2. Pro zpracování a získání klíčových bodů na obličeji je využit model MediaPipe a jeho knihovny. Konkrétně využíváme předtrénovaný model `face_landmarker_v2_with_blendshapes.task`, který slouží jako vstup pro inicializační volbu modelu. Tento model umožňuje detekci klíčových bodů na snímcích obličeje. Z těchto a dalších možností vytváříme objekt `detector`, který slouží k detekci klíčových bodů na snímcích obličeje. Při lokalizaci významných bodů je nutné počítat s tím, že může dojít k nepřesnému či chybnému určení některého z bodů. Grafický výsledek detekce klíčových bodů na obličeji je na obrázku 4.6, kde je zobrazena celá maska klíčových bodů překrytých přes původní snímek obličeje.



Obrázek 4.6: Ukázka vizualizace výsledků detekce antropometrických bodů obličeje modelu MediaPipe.

Před detekcí antropometrických bodů na jednotlivých snímcích je nezbytné snímky získat. Slouží k tomu primárně dvě pomocné funkce popsané v sekci 4.7 pro práci se soubory a adresáři. První z nich seřadí soubory v adresáři a druhá funkce následně získá absolutní cesty pro každý obrázek k dalšímu zpracování, pokud je vstupem programu pouze jedna dvojice snímků, nedochází k seřazení a je navrácena pouze absolutní cesta snímků. Poté již dochází k detekování bodů na obličeji pomocí detektoru `detector` a uložení výsledků do proměnný, pro každou skupinu, reálných a generovaných snímků, zvlášt.

Další proces spočívá v měření jednotlivých poměrů, které jsou uložené jako seznam v konstantě `FACE_PROPOTIONS`, která je získána z pomocného souboru `landmark_info.py`, který je více popsaný v sekci 4.4. Pro každé měření je spočítána vzdálenost dvou bodů v trojrozměrném prostoru (MediaPipe poskytuje pro každý bod tři souřadnice x , y a z). Výpočet vzdálenosti probíhá ve funkci `calculate_distances` a je založen na eukleidovské vzdálenosti vyjádřené vzorcem

$$distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (4.1)$$

kde $distance$ značí vypočítanou vzdálenost a x_1 , y_1 a z_1 jsou souřadnice jednoho bodu a x_2 , y_2 a z_2 bodu druhého.

Funkce, která provádí celý tento proces, nese název `measure_face_proportions`. Vrací slovník vzdáleností, kde je klíčem název měření a hodnotou je naměřená vzdálenost.

Získání hodnot vzdáleností z reálných a generovaných snímků je porovnáno pomocí funkce `compare_proportions`, kde se vypočítává rozdíl pro každou dvojici klíčů. Nad všemi absolutními hodnotami porovnání je vypočítán průměr uložený do `avg_diff`.

Program dále pokračuje funkcí `max_absolute_proportion_difference`, která provádí analýzu poměrů s největším rozdílem vzdálenosti, tedy těch měření, u kterých se proporce nejvíce liší. Prochází všechny hodnoty výsledku srovnání a vybírá tu, který má nejvyšší absolutní hodnotu, tedy i největší rozdíl ve vzdálenosti. Hodnoty jsou uloženy pro další rozbor v rámci experimentů.

Paralelně jsou získány hodnoty potřebné k výpočtu hodnoty celkového indexu obličeje (TFI). Potřebná je šířka a výška obličeje, pro přístup k této proporcím slouží konstanta `TFI`. Hodnoty jsou získány pomocí funkce `calculate_distances` a jsou uloženy do příslušných seznamů a je vypočítána hodnota TFI pomocí funkce `calculate_TFI`, také uložená do příslušného seznamu (`real_tfi` nebo `gen_tfi`). Při jakémkoliv rotaci obličeje může dojít ke zkreslení

4.6 Porovnání uměle vygenerovaných a reálných snímků obličeje

Druhým blokem pro vyhodnocení rozdílnosti mezi generovanými a reálnými snímků je využití odlehčeného (*lightweight*) rámce **Deepface** [31]. Tento rámec poskytuje širokou škálu funkcí, které umožňují přesné srovnání obličejů a analýzu jejich podobnosti, využívá knihovnu `tensorflow`. Jeho použití umožňuje přímočaré využití konkrétních funkcí s možností volby téměř libovolného modelu pro vykonání úlohy.

Program pracuje se dvěma adresáři obsahujícími reálné a generované snímků, stejně jako bylo popsáno v sekci 4.5. Každá dvojice snímků je následně porovnána pomocí vestavěné funkce rámce Deepface. Konkrétně se jedná o funkci pro ověřování obličejů `verify()`, která

slouží k ověřování snímků obličeje. Pro rozpoznávání obličeje je použit model ArcFace, který je známý svou vysokou přesností a robustností. Pro detekci obličeje je využíván detektor retinaface (odlišnost detektorů je na obrázku 3.13). Pro metriku měření podobnosti byla zvolena kosinová vzdálenost jako výchozí hodnota funkce.

Samotné volání funkce vypadá následovně

```
result = DeepFace.verify(real_image_path, gen_image_path,  
                         model_name="ArcFace",  
                         detector_backend='retinaface')
```

kde výsledek porovnání `result` poskytuje informace o podobnosti mezi obličeji na snímcích, což umožňuje detailní analýzu rozdílů mezi generovanými a reálnými obličeji. Pro program jsou zásadní následující informace z výsledku:

- `verified` (ověřeno) – hodnota vyjadřující, jestli se v porovnání jedná o stejnou osobu (`True`) nebo nikoliv (`False`)
- `distance` (vzdálenost) – hodnota vzdálenosti mezi vektory obličeje, čím nižší hodnota vzdálenosti, tím větší podobnost mezi obličejovaly rysy
- `max_threshold_to_verify` (prahová hodnota ověření) – maximální hodnota vzdálenosti pro vyhodnocení stejné identity

Pro prezentaci výsledků je využívána knihovna `matplotlib.pyplot`, která umožňuje vizualizaci dat prostřednictvím grafů. Jedním z grafických vyobrazení výsledků je **kruhový diagram** (angl. *pie chart*). Zobrazuje procentuální úspěšnost porovnání obličejů s výsečemi úspěch a neúspěch. Druhým typem grafu využívaným k prezentaci výsledků je **korelační diagram** (angl. *scatter plot*). Zobrazuje jednotlivé hodnoty vektorů vzdálenosti pro každou z dvojic. Vykreslena je i prahová hodnota (*threshold*) pro určení úspěšnosti.

4.7 Pomocné skripty a funkce

Tato sekce popisuje skripty a funkce, které pomáhají s některými procesy při tvorbě této práce. Všechny uvedené funkce a příslušné soubory jsou uloženy v adresáři `helpers`, který slouží jako úložiště pro pomocné nástroje a funkce používané v rámci projektu.

Vytvoření podmnožiny datové sady

Použitý dataset je pro tento typ práce moc velký, proto je nutné využít jeho podmnožinu. Získání této podmnožiny probíhá stochastickým výběrem napříč celým datasetem dokud není dosaženo požadované velikosti podmnožiny. Tento postup umožňuje získat reprezentativní vzorek dat pro daný úkol. Implementované funkce se nachází v souboru `create_subset.py`.

Několik implementovaných pomocných funkcí slouží k manipulaci se soubory mateřského datasetu. Hlavní funkce je `copy_random_images`, která kopíruje náhodné obrázky z daného zdroje do cílového adresáře. Pokud je zdrojem zip soubor, nejprve je extrahován do dočasného adresáře a poté jsou z něj vybrány náhodné obrázky ke kopírování. Tento proces vykonává funkce `extract_zip_to_temp`, která slouží k extrakci obsahu zip souboru do dočasného adresáře. Ve funkci `parse_arguments_create_subset` jsou zpracovány argumenty příkazové řádky a volá se hlavní funkce `copy_random_images` s nimi.

Zobrazení podmnožiny datové sady jako obrázek

Pomocná funkce pro vytvoření obrázku ve stylu mřížky pro zobrazení všech položek v podmnožině datové sady v daném adresáři. Obrázek se využívá jako ukázka subsetu pro připojení do textu práce. Implementované funkce se nachází v souboru `grid_image_subset.py`.

Po zavolání hlavní funkce `create_grid_image` se nejprve přes funkci `load_images` načtou obrázky z daného adresáře, přičemž akceptují se pouze soubory s příponami JPG, PNG a JPEG. Následně funkce `create_subplot_grid` vytváří mřížku obrázků z načtených obrázků, která se používá k vizualizaci. Nakonec je zde funkce `save_subplot_grid`, která ukládá vytvořenou mřížku obrázků do souboru. Snímky jsou seřazeny.

Vykreslení antropometrických bodů na snímek obličeje

Tato pomocná funkce slouží pro vizualizaci výsledků detekce antropometrických bodů obličeje modelu MediaPipe [25]. Část zdrojového kódu byla převzata od autorů MediaPipe a je licencována pod Apache License 2.0⁴.

První funkce, která načítá model MediaPipe a provádí detekci významných bodů na obličeji, nese název `get_image_with_landmarks`. Další částí je načtení snímku obličeje a po úspěšné operaci také uložení anotovaného obrázku. Tato funkce rovněž volá funkci `draw_landmarks_on_image`, která se stará o vizualizaci detekované mapy antropometrických bodů a vrací upravený snímek. Skript je spouštěn přes terminál s cílovou cestou k vybranému obrázku obličeje. Výstup této funkce je možné vidět na obrázku 4.6. Implementované funkce se nachází v souboru `mediapipe_draw_face_landmarks.py`.

V průběhu práce byl tento skript rozšířen o specifické zobrazení bodů, které odpovídají bodů použitým pro měření v sekci experimentů MediaPipe. Implementace je ve funkci `draw_specific_landmarks_on_image`, která ve smyčce vykresluje definované body i s popisem bodu na obrázek. Výsledek je možné vidět na obrázku 4.5.

Operace se soubory

Pro manipulaci se soubory a složkami slouží funkce v souboru `file_operations.py`. Skript využívá knihovnu `os`, která umožňuje práci s operačním systémem.

Funkce `get_sorted_files` na základě parametru `path`, označujícího cestu k cílovému adresáři, vrací seřazený seznam souborů ve vybraném adresáři.

Další funkce `get_image_paths` poskytuje absolutní cestu k souboru pro vybraný obrázek v konkrétním adresáři.

Pokud je vstupem programu dvojice obrázků namísto adresáře, vracejí obě funkce pouze cestu k obrázku. Vyhodnocení probíhá na základě podmínky `os.path.isfile(path)`.

Generování tabulky pro L^AT_EX

Pomocný skript sloužící pro generování výsledků experimentů ve formě L^AT_EX tabulky. Pro generování tabulky není použita žádná externí knihovna; používá se pouze jeden řetězec, který je postupně naplnován tak, aby vytvořil tabulku. Tvoří ji obecná struktura a poté se z poskytnutého slovníku generují jednotlivé řádky tabulky pro každou z položek. Součástí skriptu je také funkce obsahující slovník pro překlad na české alternativy klíčů, ze kterých jsou generovány konkrétní názvy položek. Druhá tabulka ve skriptu nepracuje s klíči, ale pouze generuje tabulku s poskytnutými hodnotami.

⁴<https://www.apache.org/licenses/LICENSE-2.0>

Kapitola 5

Experimenty a vyhodnocení

Experimenty byly prováděny na dvojicích reálných i uměle generovaných snímků, s využitím výpočtů v programovacím jazyce Python, kde se data zpracovávala a probíhaly s nimi další výpočty. Měření poskytly modely MediaPipe a Deepface.

5.1 Antropometrická analýza

V sekci experimentů s modelem MediaPipe dochází k vyhodnocení výsledků z antropometrické analýzy obličeje, porovnávající skutečné a umělé snímky. Po získání výsledků, které jsou popsané v sekci 4.5, program disponuje dvojicí proměnných: `max_key`, která představuje proporcí s největším rozdílem vzdálenosti k původnímu snímkmu a `max_value`, což je hodnota tohoto rozdílu. Tyto hodnoty jsou uloženy do slovníku `max_differences`. Každá položka v tomto slovníku představuje jedno měření proporcí. Kromě toho je k dispozici datová třída `MaxDifference` s atributy `count`, což značí počet výskytů získaných postupným navýšováním pro každý prvek ve stejném měření, a `value`, což je součet všech rozdílů pro dané měření. Pro každé měření je nad tímto součtem rozdílů proveden aritmetický průměr, aby se získala průměrná hodnota rozdílu.

Měřená proporce	Počet výskytů	Průměrný rozdíl
výška obličeje (tr-gn)	32	0,0218
šířka úst (ch-ch)	35	0,0179
speciální výška obličeje vpravo (en-gn)	4	0,0153
šířka mezi oči vnější (ex-ex)	27	0,0155
výška dolní třetiny obličeje (sto-gn)	9	0,0174
šířka čelisti (go-go)	49	0,0234
speciální výška horní části obličeje (g-sn)	5	0,0147
výška horní části obličeje (n-sto)	8	0,0181
šířka obličeje (zy-zy)	50	0,0216
výška dolní části obličeje (sn-gn)	8	0,0163
šířka nosu (al-al)	2	0,0134
morfologická výška obličeje (n-gn)	5	0,0134
šířka mezi oči vnitřní (en-en)	4	0,0140
speciální výška obličeje vlevo (en-gn)	11	0,0177
šířka pravého oka (ex-en)	1	0,0066

Tabulka 5.1: Tabulka výsledků antropometrické analýzy.

Tabulka 5.1 ukazuje naměřené hodnoty experimentu pro 250 dvojic reálných a uměle vygenerovaných snímků obličeje.

Proporce s nejvyšším počtem výskytů, pokud jde o největší rozdíly ve všech měřených proporcích pro jednu dvojici, jsou šířka obličeje (zy-zy), následovaná šírkou čelisti (go-go), u které se antropometrické body nacházejí na podobných místech jako u celkové šírky obličeje. Další nejčastěji se vyskytující proporcí je šířka úst (ch-ch). První tři pozice z hlediska četnosti výskytu obsadily proporce šírky obličeje. Na čtvrtém místě je výška obličeje (tr-gn). Pro tyto proporce lze říct, že se nejvíce liší při generování umělých snímků.

Proporce, které nikdy nebyly největším rozdílem vzdáleností v dvojici a jsou tedy nejméně časté, jsou výška horního rtu (ls-sto) a výška dolního rtu (sto-li). Dále šířka levého oka (ex-en), pokud jde o šířku pravého oka (ex-en), vyskytuje se pouze jednou. Výška nosu (n-sn) je další hodnotou, která se v měření nikdy nevyskytla, šířka nosu (al-al) byla zaznamenána pouze dvakrát. Všechny tyto proporce se jen minimálně lišily od původního snímku. Zmíněné proporce patří k těm menším na celém obličeji.

Při zaměření na průměrný rozdíl pro každou z proporcí z tabulky 5.1, nejvyšší průměrný rozdíl je zaznamenán u proporce šírky čelisti (go-go), následovaný proporcí šírky obličeje (zy-zy), která kopíruje první dvě místa ve zkoumání, avšak v opačném pořadí. Třetí proporcí s rozdílem přesahujícím hodnotu 0,02 je výška obličeje (tr-gn), která je zároveň posledním takovým případem.

Při nahlédnutí, jak velkou část tyto odchylky tvoří je možné použít hodnoty z tabulky 5.2 a vydělit je s rozdíly jednotlivých proporcí. Z výpočtu vyplývá, že rozdíl šírky obličeje tvoří přibližně 3,9 % průměrné vzdálenosti této proporce, zatímco rozdíl výšky obličeje představuje 4,35 %.

Kromě maximálního průměrného rozdílu mezi proporcemi je pro každou porovnanou dvojici zachována také průměrná hodnota všech výsledků porovnání proporcí. Aritmetický průměr jednotlivých průměrných hodnot pro všechny porovnávané dvojice se rovná hodnotě rozdílu vzdáleností 0,00633. Tato hodnota představuje průměrný rozdíl napříč všemi porovnanými proporcemi.

Celkový index obličeje

Dalším experimentem bylo provedení měření a výpočtu celkové indexu obličeje vycházející ze vzorce (2.1). Pro jeho výpočet je potřebná morfologická šířka a výška obličeje. Následně se počítá procentuální poměr mezi těmito vzdálenostmi. Výsledky pro reálné, uměle vygenerované snímkы a jejich spojení jsou zobrazeny v tabulce 5.2. Všechny hodnoty jsou aritmetické průměry všech hodnot z dané kategorie.

Parametr	Reálný snímek	Umělý snímek	Kombinace
MFH	0,50058	0,49969	0,50014
MFB	0,54835	0,55609	0,55222
TFI	91,36994	89,91802	90,64398

Tabulka 5.2: Tabulka zobrazující výsledky výpočtu celkového indexu obličeje (TFI), který vyjadřuje poměr mezi výškou (MFH) a šírkou (MFB) obličeje. U měření hodnoty reprezentují vzdálenost mezi dvěma antropometrickými body. U TFI představují hodnoty procentuální poměr.

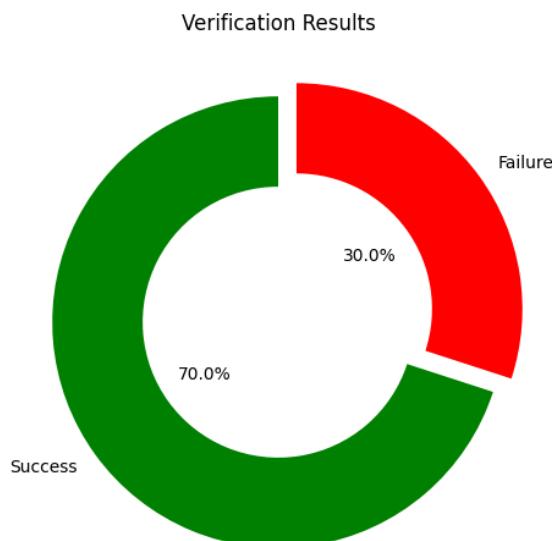
Rozdíl mezi vzdáleností vektorů reálného a umělého snímku je pro výšku obličeje 0,00089 a pro šířku 0,00774, což je třetina naměřeného rozdílu v předchozím experimentu, je zde nutné brát v potaz, že předchozí experiment vyhodnocoval pouze výkyvy v rozdílech vzdáleností, ale zde se počítá se všemi naměřenými hodnotami.

Vypočítané obličejové poměry se mezi reálnými a umělými snímky zásadně neliší a v obou případech spadají do skupiny *leptoprosopic* [15] ($88,0 < \text{TFI} < 92,9$), to se týče i celkového průměru.

5.2 Srovnávací analýza

Experimenty provedené v této sekci vycházejí z rámce Deepface a využívají návratovou hodnotu funkce `verify`, popsanou v sekci 4.6. Zkoumání probíhá na dvou úrovních: první úroveň analyzuje, kolik z výsledků porovnání dvojic reálných a uměle vytvořených snímků obličeje je úspěšných, tedy kolik z nich označuje dvojice za identické osoby. Druhá úroveň se detailněji zaměřuje na konkrétní data každé dvojice a jak velký se u nich vyskytl rozdíl.

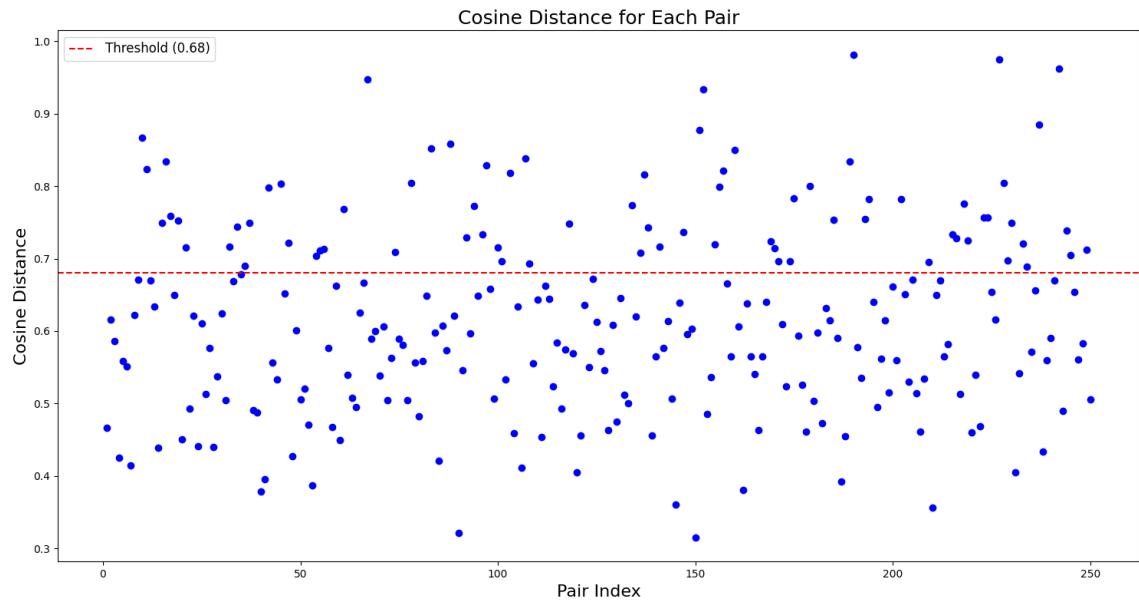
První experiment se zaměřuje na porovnání identit dvojic reálných a uměle vygenerovaných snímků na základě atributu `verified` proměnné `result`. Tento atribut může nabývat hodnoty `True`, což indikuje úspěšné porovnání, kde identita obou dvojic je totožná, nebo `False`, což naznačuje opak. Výsledky experimentu provedeného na 250 dvojicích jsou zobrazeny v grafu 5.1, který ukazuje procentuální zastoupení obou hodnot. Tento graf vyjadřuje procentuální počet úspěšných srovnání identit.



Obrázek 5.1: Kruhový diagram pro zobrazení úspěšnosti porovnání obličejů rámcem Deepface. Zelená část značí shodu v porovnání (*Success*) a červená značí neúspěch (*Failure*).

Z diagramu 5.1 je možné vidět, že 70 % (175) dvojic bylo vydobnocoeno jako úspěšné a byla zde zjištěna shoda v podobnosti. Naopak 30 % dvojic nebylo vydobnocoeno jako podobné a hodnota vzdálenosti vektorů *distance* u nich přesáhla prah podobnosti.

Druhá část experimentu prezentuje vypočtenou vzdálenost *distance* mezi dvěma vektory obličeje, kde nižší hodnota indikuje větší podobnost mezi dvojicemi. Vzdálenost je zobrazena pro všech 250 dvojic z provedeného experimentu v grafu 5.2. V grafu figuruje i červená přerušovaná čára, představující prahovou hodnotu *threshold*, která byla použita k rozhodnutí o podobnosti v předchozím experimentu. Všechny hodnoty vzdáleností pod touto čarou jsou vydobnocoeny jako shoda v porovnání. Prahová hodnota pro tento experiment byla 0,68.

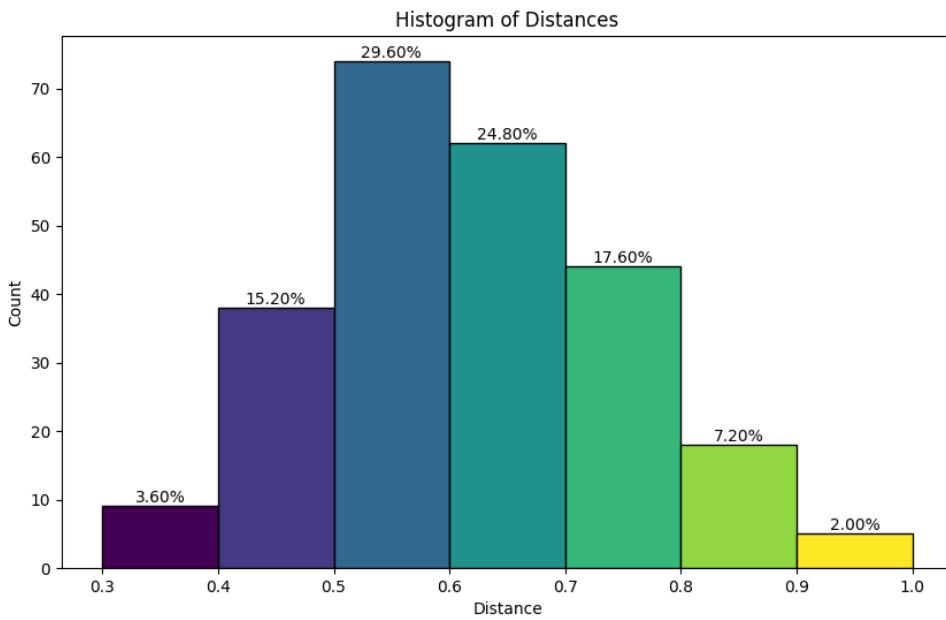


Obrázek 5.2: Korelační diagram pro zobrazení vzdálenosti vektorů dvojcí snímků obličeju. Červená přerušovaná čára zastupuje prahovou hodnotu (*threshold*), rozhodující o úspěšnosti porovnání.

Diagram 5.2 zobrazuje grafické znázornění výsledků vydobnocoujících podobnost dvojcí. Při bližším zkoumaní hodnot determinující úspěšnost při předchozí části experimentu bylo zjištěno, že průměrná hodnota vzdálenosti *distance* byla 0,61577 a medián hodnot byl 0,60772. Obě hodnoty se nacházejí pod prahovou hodnotou 0,68, podporující vyšší úspěšnost s předchozí částí experimentu.

Nejnižší naměřená hodnota vzdálenosti byla 0,3153, zaznamenaná u páru označeného číslem 150. Nejvyšší vypočítaná hodnota byla 0,9815 u páru 190, s nejmenší podobností. Více o extrémech v hodnotách vzdáleností v sekci 5.3.

Nejvíce dvojcí se vyskytovalo v rozmezí od 0,5 do 0,6 – konkrétně se jednalo o 74 páru, což představuje 29,6 % všech dvojcí. Nejméně zastoupený byl interval 0,9 až 1, kde bylo zaznamenáno pouze 5 výskytů, tvořící 2 % všech vzdáleností. Procentuální zastoupení vzdáleností pro všechny intervaly jsou na obrázku 5.3.



Obrázek 5.3: Procentuální zastoupení na intervalech pro všechny naměřené vzdálenosti.

Tento experiment přináší podrobný pohled na úspěšnost porovnání identit mezi dvojicemi reálných a uměle vytvořených snímků obličeje, stejně jako na rozdíly v jejich vektorových reprezentacích s podrobným pohledem na tyto hodnoty. Současně připravuje půdu pro kombinované experimenty spojením s naměřenými hodnotami antropometrické analýzy.

5.3 Kombinovaná analýza

Experimenty v této sekci budou pracovat s výsledky z předchozích experimentů a budou kombinovat oba použité modely a jejich výstupy, pro podrobnější zkoumání. Hlavním účelem bude zkoumat podobnost dat mezi modely a zjistit, zda extrémní hodnoty u jednoho modelu kopírují hodnoty u modelu druhého. Konkrétně budou analyzovány dvojice s nejvyššími a nejnižšími hodnotami vzdálenosti podle rámce Deepface. U těchto nově vytvořených podskupin bude provedena antropometrická analýza modelem MediaPipe a vyhodnotí se, zda jsou zde také data nakloněné k extrému.

Dvojice s nejvyšší vzdáleností vektorů

Výskytů s hodnotou vzdálenosti vektorů obličeje vyšší než 0,9 bylo pouze 5 z celkového počtu 250 zkoumaných dvojic. Tyto páry měly nejvyšší srovnávací hodnoty v rámci celé studie, což naznačuje, že obličeje by měly být nejvíce odlišné. Tabulka 5.3 ukazuje konkrétní hodnoty vzdáleností mezi těmito páry, získané pomocí funkce rámce Deepface rozšířené a průměrnou hodnotu rozdílů vzdáleností modelu MediaPipe.

Označení snímku	Vzdálenost Deepface	Průměrná vzdálenost MediaPipe
067	0,9469	0,0104
152	0,9335	0,0087
190	0,9815	0,0067
227	0,9753	0,0103
242	0,9625	0,0079

Tabulka 5.3: Tabulku zahrnující nejvyšší spočítané vzdálenosti vektorů pro dvojice snímků.

Hodnoty antropometrického měření u této skupiny jsou vyšší než obecný průměr 0,00633 všech dvojic. Průměrná hodnota u modelu MediaPipe v této skupině činí 0,0088. Vyšší vzdálenost byla tedy zjištěna nejen při analýze pomocí modelu Deepface, ale také při analýze modelem MediaPipe.

Dvojice s nejnižší vzdáleností vektorů

Obdobně jako předchozí podsekce se zde pracuje s hodnotami vzdáleností od rámce Deepface, zde se ale jedná o ty nejnižší, tedy ty páry, které si byly nejvíce podobné. Pracuje se zde s dvojicemi s hodnotou nižší než 0,4. Jedná se o 9 z celkového počtu 250 zkoumaných dvojic, tvořící 3,6 % výsledků. Výsledky vzdáleností jsou zobrazeny v tabulce 5.4, společně s hodnotami měření modelu MediaPipe.

Označení snímku	Vzdálenost Deepface	Průměrná vzdálenost MediaPipe
040	0,3788	0,0053
041	0,3952	0,0049
053	0,3869	0,0057
090	0,3210	0,0055
145	0,3606	0,0051
150	0,3153	0,0062
162	0,3805	0,0095
187	0,3921	0,0044
210	0,3567	0,0052

Tabulka 5.4: Tabulku zahrnující nejnižší spočítané vzdálenosti vektorů pro dvojice snímků.

U skupiny s nejvyšší podobností a nejnižší vzdáleností byla spočítána průměrná hodnota pro model MediaPipe ve výši 0,0057, což odpovídá výsledkům získaným od rámce Deepface. Průměrná hodnota vzdáleností jednotlivých proporcí je výrazně nižší než celkový průměr modelu MediaPipe, který je 0,0063.

Experimenty se zaměřily na zkoumání extrémních hodnot (vysoké a nízké) obou modelů, aby zjistily, zda jsou v souladu. V obou experimentech bylo prokázáno, že vyšší hodnoty získané z jednoho modelu vedou k vyšším hodnotám z modelu druhého modelu. Totéž platí i pro nižší hodnoty; pokud model první generuje nižší hodnoty, model druhý bude vykazovat podobný trend. To znamená, že výsledky experimentů jsou vzájemně konzistentní a podporují hypotézu, že oba modely sdílejí podobné trendy v měření a analýzách.

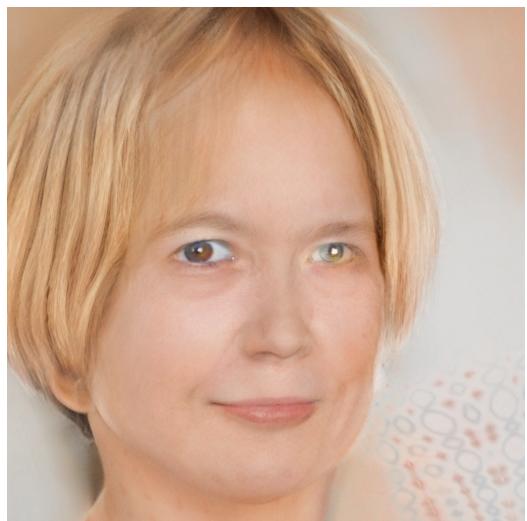
5.4 Obecný pohled na dvojice

V této sekci bude představen autorův osobní pohled na rozdílnost reálných a uměle vygenerovaných snímků obličeje. Zkoumány budou dvě dvojice, každá z opačného spektra skupin minimálních a maximálních vzdáleností vektorů uvedených v sekci 5.3. Ke konkrétním dvojicím bude doplněn komentář o rozdílech mezi obličeji, včetně upozornění na případně vzniklé artefakty a jejich rozsáhlosti.

První zkoumanou dvojicí budou snímky obličeje s pořadovým číslem 242, patřící do skupiny s nejvyšším rozdílem vzdáleností vektorů a z technické stránky tedy jedna z nejvíce lišících se dvojic. Oba snímky jsou na obrázku 5.4.



(a) Původní snímek obličeje

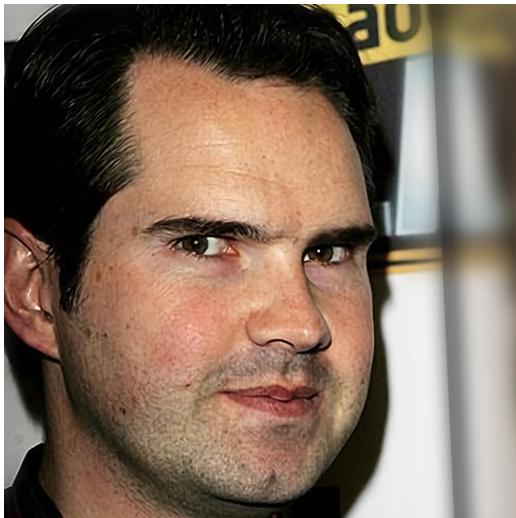


(b) Umělý vygenerovaný snímek obličeje

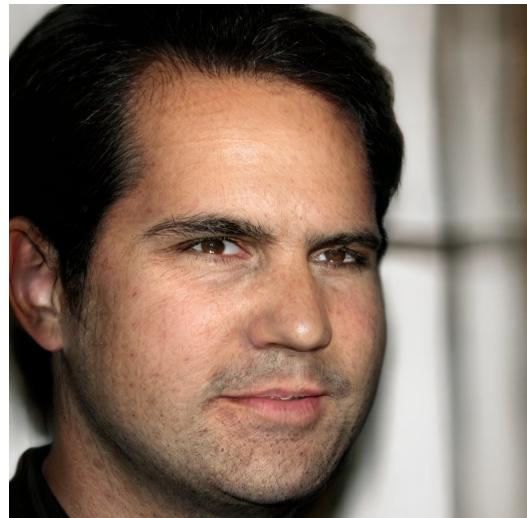
Obrázek 5.4: Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3 s vysokou hodnotou vzdálenosti mezi vektory obličeje.

O dvojici lze říci, že se na první pohled jeví jako velmi odlišná. Jeden ze spojujících prvků jsou vlasy, kterým zůstal stejný odstín jako na původním snímku. Co se týče proporcí – šířka obličeje je výrazně větší u generovaného snímku, to stejně lze říct i o šířce čelisti. Rozdíl ve výšce již tolik vidět nelze. Výrazný rozdíl lze sledovat dále u úst, které disponují úplně odlišným tvarem, podobná situace nastala i u očí, kde došlo i k výrazné změně barvy u obou očí a k odlišnému natočení pohledu. Artefaktem vystupujícím na vygenerovaném snímku je ucho, které je položené pod vlasy na levé straně tváře. Tváře ztratily lehce načervenalou barvu, ta se proměnila na více neutrální, korespondující se zbytkem obličeje.

Druhou zkoumanou dvojicí budou snímky obličeje s pořadovým číslem 90, patřící naopak do skupiny s nejnižším rozdílem vzdáleností vektorů. Reálný i vygenerovaný snímek jsou na obrázku 5.5.



(a) Původní snímek obličeje



(b) Umělý vygenerovaný snímek obličeje

Obrázek 5.5: Porovnání reálného a vygenerovaného snímku obličeje modelem StyleGAN3 s nízkou hodnotou vzdálenosti mezi vektory obličeje.

Podobně jako u předchozí dvojice zde dochází k výrazným změnám u očí a úst jedince. Oči jsou výrazně menší a ústa jsou naopak mírně širší. Šířka a výška obličeje zůstává v tomto případě poměrně podobná původnímu snímku. Kromě specifické textury obličeje, pravděpodobně způsobené nižší kvalitou původní fotografie, nejsou přítomny žádné další anomálie nebo artefakty. Stejně jako u předchozí dvojice vymizela při generování lehká červená barva pokrývající tvář obličeje, která se přetvořila na více neutrální odstín pokrývající zbytek obličeje. Nos u dvojice se výrazně změnil – stal se kulatějším a méně širokým než na původní fotografii.

Z prozkoumaných dvojic snímků je vidět trend v odlišnosti rozměrů a vzhledu očí a úst. Dále se u obou příkladů měnil barevný profil tváře na více neutrální. U dvojice s vyšší vzdáleností vektorů obličeje byl viditelnější rozdíl v šířce obličeje a čelisti, jedná se o dvě nejvíce lišící se proporce, tedy i o faktor, který nejvíce rozhodoval o neúspěšném vyhodnocení podobnosti dvojice.

Kapitola 6

Závěr

Tato práce se zaměřuje na dvě hlavní oblasti: detekci a zkoumání antropometrických bodů na obličeji a generování uměle vytvořených snímků obličeje pomocí sítí GAN. Došlo k prozkoumání toho, které antropometrické body jsou na obličeji významné a jaké proporce lze měřit a analyzovat. Byl vybrán model pro generování snímků obličeje, který využívá reálný snímek jako zdroj pro generování.

V rámci antropometrické analýzy bylo nahlíženo na významné proporce na obličeji jednotlivců a z kterých antropometrických bodů se skládají. Proporce zahrnují různé rozměry obličeje, jako je šířka a výška obličeje, šířka oka, vzdálenost očí od sebe, výška nosu, šířka úst a šířka čelisti. Kromě toho se měří další klíčové oblasti, například délka nosu, vzdálenost mezi nosem a ústy, nebo výška rtů.

Při výběru modelu pro generování umělých snímků obličeje bylo prozkoumáno několik možností, které byly vyhodnoceny z hlediska přesnosti, kvality a schopnosti generovat realistické snímky. Jako nejhodnější a nejpokročilejší model byl shledán model StyleGAN3, který se vyznačuje pokročilými schopnostmi v oblasti generování detailních a realistických obličejů. V práci zastupuje roli pro dotvoření dvojic reálných snímků podmnožiny datové sady CelebAMask-HQ.

V práci se podařilo vyhodnotit dvojice reálných a umělých snímků obličeje dvěma způsoby. Prvním byla antropometrická analýza modelem MediaPipe, která poskytovala mapu antropometrických bodů na obličeji. Byly měřeny jednotlivé definované proporce a porovnány v rámci dvojic snímků. Následně byly vybrány proporce, které se u dvojic lišily nejvíce. Největší rozdíly byly zaznamenány v šířce obličeje, čelisti a úst, stejně jako ve výšce obličeje. Naopak menší rozdíly byly zjištěny ve výšce horního a dolního rtu nebo šířce oka. Celkový poměr obličeje u reálných snímků dosáhl 91,37 % a u umělých snímků 89,92 %. Tato rozdílná hodnota je relativně malá a naznačuje lehkou odlišnost.

Druhý způsob zastupoval rámec Deepface, který poskytoval srovnání dvojic na základě výpočtu jejich kosinové vzdálenosti. Z 250 zkoumaných dvojic bylo 175 vyhodnoceno jako podobných, tvořící 70 %. Průměrný hodnota vzdálenosti byla 0,6158, přičemž prahová hodnota pro rozhodnutí o podobnosti byla 0,68. Jakákoli hodnota pod touto mezí byla považována za podobný výsledek.

Z pohledu dalšího vývoje dává smysl prozkoumat i oblast difúzních modelů pro generování obličeje a provést antropometrickou analýzu i zde. Difúzní modely jsou dnes již velice známé a jejich popularita stále roste i v široké veřejnosti.

Literatura

- [1] AI, E. of. Advanced neural network techniques. *Elements of AI* [online]. MinnaLearn, 21. května 2018 [cit. 2024-02-27]. Dostupné z: <https://course.elementsofai.com/5/3>.
- [2] CHOI, Y., CHOI, M., KIM, M., HA, J.-W., KIM, S. et al. *StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation*. 2018 [cit. 2024-02-21].
- [3] CORPORATION, I. What are neural networks? *IBM* [online]. 16. prosince 2021 [cit. 2023-12-05]. Dostupné z: <https://www.ibm.com/topics/neural-networks#:~:text=Neural%20networks%2C%20also%20known%20as,neurons%20signal%20to%20one%20another>.
- [4] COUNCIL, N. R. *Biometric Recognition: Challenges and Opportunities*. 1. vyd. Washington, DC: The National Academies Press, 2010 [cit. 2024-02-19]. ISBN 978-0-309-14207-6. Dostupné z: <https://nap.nationalacademies.org/catalog/12720/biometric-recognition-challenges-and-opportunities>.
- [5] DECARLO, D., METAXAS, D. a STONE, M. An Anthropometric Face Model using Variational Techniques. In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*. New York, New York, USA: ACM Press, 1998, s. 67–74 [cit. 2024-04-02]. DOI: 10.1145/280814.280823. ISBN 0897919998. Dostupné z: <http://portal.acm.org/citation.cfm?doid=280814.280823>.
- [6] DENG, J., GUO, J., YANG, J., XUE, N., KOTSIDA, I. et al. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Institute of Electrical and Electronics Engineers (IEEE). říjen 2022, sv. 44, č. 10, s. 5962–5979, [cit. 2024-03-19]. DOI: 10.1109/tpami.2021.3087709. ISSN 1939-3539. Dostupné z: <http://dx.doi.org/10.1109/TPAMI.2021.3087709>.
- [7] GMBH, R. *A comprehensive survey on semantic facial attribute editing using generative adversarial networks*. 2023 [cit. 2023-12-27]. Dostupné z: https://www.researchgate.net/figure/Elegants-architecture-183_fig7_360804375.
- [8] GMBH, R. *Segmentation Guided Image-to-Image Translation with Adversarial Networks*. 2023 [cit. 2023-12-27]. Dostupné z: https://www.researchgate.net/figure/Illustration-of-SGGAN-SGGAN-contains-a-a-generator-G-b-a-discriminator-with_fig2_330212752.
- [9] GOODFELLOW, I. J., POUGET ABADIE, J., MIRZA, M., XU, B., WARDE FARLEY, D. et al. *Generative Adversarial Networks*. 2014 [cit. 2023-12-23]. Dostupné z: <https://doi.org/10.48550/arXiv.1406.2661>.

- [10] GOOGLE. Overview of GAN Structure. *Google* [online]. Google Developers, 20. listopadu 2019. Aktualizováno 18.7.2022 [cit. 2023-12-24]. Dostupné z: https://developers.google.com/machine-learning/gan/gan_structure.
- [11] GUPTA, S. The 7 Most Common Machine Learning Loss Functions. *Built In* [online]. Built In, 14. dubna 2022 [cit. 2024-02-27]. Dostupné z: <https://builtin.com/machine-learning/common-loss-functions>.
- [12] HARDESTY, L. Explained: Neural networks. *MIT News / Massachusetts Institute of Technology* [online]. MIT News Office, 14. dubna 2017 [cit. 2023-12-05]. Dostupné z: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.
- [13] HU, L. Generative AI and Future. *Medium* [online]. Towards AI, 17. července 2022 [cit. 2023-12-23]. Dostupné z: <https://pub.towardsai.net/generative-ai-and-future-c3b1695876f2>.
- [14] HUGH, C. *Encyclopædia Britannica Eleventh Edition*. 2. vyd. United States: Horace Everett Hooper, 1910-1911 [cit. 2023-12-02].
- [15] JEREMIC, D., KOCIC, S., VULOVIC, M., SAZDANOVIC, M., SAZDANOVIC, P. et al. Anthropometric study of the facial index in the population of central Serbia. *Archives of Biological Sciences*. Leden 2013, sv. 65, č. 3, s. 1163–1168, [cit. 2024-04-14]. DOI: 10.2298/ABS1303163J. ISSN 0354-4664. Dostupné z: <https://doiserbia.nb.rs/Article.aspx?ID=0354-46641303163J>.
- [16] KARRAS, T., AITTALA, M., LAINE, S., HÄRKÖNEN, E., HELLSTEN, J. et al. Alias-free generative adversarial networks. *Advances in neural information processing systems*. 2021, sv. 34, s. 852–863, [cit. 2024-03-24].
- [17] KARRAS, T., LAINE, S. a AILA, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *CoRR*. 2018, abs/1812.04948. Dostupné z: <http://arxiv.org/abs/1812.04948>.
- [18] KARRAS, T., LAINE, S., AITTALA, M., HELLSTEN, J., LEHTINEN, J. et al. *Analyzing and Improving the Image Quality of StyleGAN*. 2020 [cit. 2024-03-24].
- [19] KAZEMI, V. a SULLIVAN, J. One millisecond face alignment with an ensemble of regression trees. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2014, s. 1867–1874, [cit. 2023-12-29]. DOI: 10.1109/CVPR.2014.241. Dostupné z: <https://api.semanticscholar.org/CorpusID:2031947>.
- [20] KOLF, J. N., RIEBER, T., ELLIESEN, J., BOUTROS, F., KUIJPER, A. et al. *Identity-driven Three-Player Generative Adversarial Network for Synthetic-based Face Recognition*. 2023 [cit. 2024-02-19].
- [21] KOMLOS, J. Anthropometric history: an overview of a quarter century of research. *Anthropologischer Anzeiger* [online]. 1. vyd. Stuttgart, Germany: Schweizerbart Science Publishers. Prosinec 2009, sv. 67, č. 4, s. 341–356, [cit. 2023-12-02]. DOI: 10.1127/0003-5548/2009/0027. Dostupné z: <http://dx.doi.org/10.1127/0003-5548/2009/0027>.

- [22] LEE, C.-H., LIU, Z., WU, L. a LUO, P. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, s. 5549–5558 [cit. 2023-12-27].
- [23] LIU, M., DING, Y., XIA, M., LIU, X., DING, E. et al. *STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing*. 2019 [cit. 2024-04-19].
- [24] LIU, Z., LUO, P., WANG, X. a TANG, X. Deep Learning Face Attributes in the Wild. *CoRR*. 2014, abs/1411.7766, [cit. 2024-03-09]. DOI: <https://doi.org/10.48550/arXiv.1411.7766>. Dostupné z: <http://arxiv.org/abs/1411.7766>.
- [25] LUGARESI, C., TANG, J., NASH, H., MCCLANAHAN, C., UBOWEJA, E. et al. MediaPipe: A Framework for Perceiving and Processing Reality. In: *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*. 2019 [cit. 2024-03-31]. Dostupné z: https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf.
- [26] LUGARESI, C., TANG, J., NASH, H., MCCLANAHAN, C., UBOWEJA, E. et al. MediaPipe: A Framework for Building Perception Pipelines. *CoRR*. 2019, abs/1906.08172, [cit. 2024-04-09]. Dostupné z: <http://arxiv.org/abs/1906.08172>.
- [27] NGUYEN, H. V. a BAI, L. Cosine Similarity Metric Learning for Face Verification. In: KIMMEL, R., KLETTE, R. a SUGIMOTO, A., ed. *Computer Vision – ACCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, s. 709–720 [cit. 2024-04-01]. ISBN 978-3-642-19309-5.
- [28] OPTICAL, B. Rees-Fairbanks Facial Gauge. *Birmingham Optical* [online]. Prosinec 2023 [cit. 2023-12-03]. Dostupné z: <https://www.birminghamoptical.co.uk/product/rees-fairbanks-facial-gauge/>.
- [29] ROSEBROCK, A. Facial landmarks with dlib, OpenCV, and Python. *PyImageSearch* [online]. PyImageSearch, 3. dubna 2017. Aktualizováno 3.7.2021 [cit. 2023-12-28]. Dostupné z: <https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>.
- [30] SAGONAS, C., ANTONAKOS, E., TZIMIROPOULOS, G., ZAFEIRIOU, S. a PANTIC, M. 300 Faces In-The-Wild Challenge: database and results. *Image and Vision Computing*. 2016, sv. 47, s. 3–18, [cit. 2023-12-29]. DOI: <https://doi.org/10.1016/j.imavis.2016.01.002>. ISSN 0262-8856. 300-W, the First Automatic Facial Landmark Detection in-the-Wild Challenge. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0262885616000147>.
- [31] SERENGIL, S. I. a OZPINAR, A. LightFace: A Hybrid Deep Face Recognition Framework. In: IEEE. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. 2020, s. 23–27 [cit. 2024-03-17]. DOI: 10.1109/ASYU50717.2020.9259802. Dostupné z: <https://doi.org/10.1109/ASYU50717.2020.9259802>.
- [32] SHARMA, S. What the Hell is Perceptron? *Medium* [online]. Towards Data Science, 9. září 2017 [cit. 2023-12-04]. Dostupné z: <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>.

- [33] TECH AI MATH. Neural Networks: Unveiling the Architecture of Intelligence. *Medium* [online]. Artificial Intelligence in Plain English, 27. srpna 2023 [cit. 2024-02-19]. Dostupné z: <https://ai.plainenglish.io/neural-networks-unveiling-the-architecture-of-intelligence-2a7ca9969f73>.
- [34] WANG, L., CHEN, W., YANG, W., BI, F. a YU, F. R. A State-of-the-Art Review on Image Synthesis With Generative Adversarial Networks. *IEEE Access*. 2020, sv. 8, s. 63514–63537, [cit. 2023-12-27]. DOI: 10.1109/ACCESS.2020.2982224.
- [35] WEISSTEIN, E. W. Convolution. *MathWorld—A Wolfram Web Resource*. Wolfram Research, Inc. 2003, [cit. 2024-05-06]. Dostupné z: <https://mathworld.wolfram.com/Convolution.html>.
- [36] ZACHAROPOULOS, G. V., MANIOS, A., KAU, C. H., VELAGRAKIS, G., TZANAKAKIS, G. N. et al. Anthropometric Analysis of the Face. *Journal of Craniofacial Surgery*. 2016, sv. 27, č. 1, s. e71–e75, [cit. 2024-04-02]. DOI: 10.1097/SCS.0000000000002231. ISSN 1049-2275. Dostupné z: <https://journals.lww.com/00001665-201601000-00101>.