



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

VLIV VÝBĚRU ZDROJOVÉHO SNÍMKU NA KVALITU VÝSLEDNÉHO DEEPFAKE

AUTOŘI PRÁCE

Bc. ADAM ŠMEHÝL

Bc. PETR BALOK

BRNO 2023

Abstrakt

Tato práce se zabývá vlivem výběru zdrojového snímku na kvalitu výsledných deepfake. Pro tvorbu deepfaků využíváme moderní nástroj FaceFusion. Výsledky analyzujeme nástroji pro rozpoznání obličejů CompreFace a MagFace. Při výběru nevhodných snímků se zaměřujeme na snímky se zakrytím obličeje, které dělíme na tři typy zakrytí obličeje: vlasy, brýle a jiné překrytí, jako je ruka nebo oblečení, a analyzujeme jejich dopad na výsledný deepfake.

Z výsledků vyplývá, že i s nekvalitními zdrojovými snímky lze vytvořit přesvědčivé deepfakes, což má významné důsledky pro autenticitu digitálního obsahu. Použití nástrojů CompreFace a MagFace pro rozpoznání obličejů odhaluje různé úrovně efektivity v závislosti na typu zakrytí obličeje. Tato práce přináší důležité poznatky pro oblast digitální bezpečnosti a autenticity, poukazuje na potřebu dalšího výzkumu a zdůrazňuje možná rizika spojená s deepfakes v digitálním prostředí.

Klíčová slova

deepfake, FaceFusion, rozpoznání obličejů, GFPGAN, InsightFace, CompreFace, MagFace, CelebA, CelebA-HQ, bezpečnost digitálního obsahu

Obsah

1	Úvod	2
2	Příprava	3
2.1	Dataset CelebA	3
2.1.1	Identifikace a třízení snímků	3
2.1.2	Výběr snímků	4
2.1.3	Transformace vybraných snímků	4
2.2	Deepfakes	5
2.2.1	FaceFusion	5
2.3	Face recognition	7
2.3.1	CompreFace	7
2.3.2	MagFace	9
3	Experimentování a vyhodnocení	10
3.1	Tvorba deepfakes	10
3.2	Rozpoznání obličejů prvním nástrojem	10
3.3	Vyhodnocení dat	12
3.4	Zkvalitnění dat použitím dalšího nástroje	14
3.5	Vyhodnocení výstupu	15
4	Závěr	16
	Literatura	17

Kapitola 1

Úvod

V současné době, kdy digitální technologie dramaticky přetváří naše vnímání reality, stojíme před nezbytností porozumět a analyzovat důsledky těchto technologií. Jedním z nejzásadnějších vývojů v oblasti digitální manipulace je technologie deepfake, která umožňuje vytvářet přesvědčivé a těžko odhalitelné manipulace s obličejobými rysy v digitálním obsahu. Tato práce se zabývá vlivem výběru zdrojového snímku na kvalitu deepfake, přičemž klade důraz na princip "Garbage in => Garbage out" v kontextu algoritmů pro zpracování snímků.

V našem projektu jsme se zaměřili na použití nejnovějších nástrojů pro generování deepfaků, jako je FaceFusion, který využívá moderních modelů pro rozpoznání obličejů, jako je InsightFace, společně s algoritmy pro zlepšení obrazu, jako je GFGAN. Pomocí nástrojů pro rozpoznávaní obličejů jsme zkoumali, jak různé typy zakrytí obličeje (vlasy v obličeji, brýle a jiné překrytí) ovlivňují kvalitu a realističnost výsledného deepfake.

Specificky jsme se zaměřili na tři typy překrytí:

1. Vlasy v obličeji, kde vlasy osoby na snímku znatelně překrývají část obličeje, jako jsou oči, ústa nebo nos.
2. Brýle, kde osoba na snímku nosí brýle
3. Jiné překrytí, kde obličeje osoby je zakrytý předměty, rukou nebo oblečením.

Kapitola 2

Příprava

Pro zhotovení deepfaků a následné vyhodnocení kvality výsledků budeme potřebovat tři hlavní komponenty. Tyto komponenty zahrnují vstupní snímky (dataset), nástroj pro tvorbu deepfakes a nástroj pro rozpoznání obličejů (face recognition). Vybraný dataset bude nutné protřídit a transformovat do vhodné podoby, například upravit rozlišení snímků nebo upravit formát souborů. Je také nutné vybrat vhodné nástroje a tyto nástroje správně zprovoznit.

2.1 Dataset CelebA

Pro práci na snímcích obličejů nám byl doporučen pravděpodobně největší a nejpoužívanější dataset CelebA. Dataset CelebA, publikovaný v roce 2015 výzkumníky z Čínské univerzity v Hong Kongu, obsahuje více než 200 tisíc snímků celebrit z internetu. Obsahuje soubory s anotacemi atributů (např. nálada, oblečení, stáří) a pozicemi pěti bodů na obličeji (oči, nos, koutky úst). Později byly zveřejněny i informace o identifikaci osob, které umožňují identifikovat snímkы identické osoby. Tento dataset je volně dostupný ke stažení pro akademické a nekomerční užití. Nicméně tento původní dataset obsahuje pouze snímky o velikosti 218 x 178 pixelů. Předpokládá se, že snímkы o takto nízkém rozlišení nebudou za žádných okolností poskytovat dobrý výsledek a taktéž se nejedná o rozlišení vhodné pro práci na projektech v roce 2023. Naštěstí byl roce 2020 zveřejněn lepší dataset CelebA-HQ [2], jenž zahrnuje 30 tisíc vybraných snímků z původního datasetu, tentokrát ale o rozlišení 1024 x 1024 pixelů. V tomto projektu pracujeme s datasetem CelebA-HQ.

2.1.1 Identifikace a třízení snímků

Dataset CelebA-HQ obsahuje složku snímků pojmenovaných typu 0.jpg až 29999.jpg. Abychom mohli porovnávat snímkы obličeje téže osoby, je nutné znát, které snímkы patří jedné osobě. Protože dataset CelebA-HQ obvykle neobsahuje přímé přiřazení identit ke snímkům, je nutné provést identifikaci osob jinak. Máme k dispozici pouze soubor, který mapuje snímkы z datasetu CelebA-HQ na původní dataset CelebA (CelebA-HQ_name, CelebA_name). K tomuto původnímu datasetu máme již mapování identit (CelebA_name, identity_id), což nám umožňuje vytvořit potřebný soubor pro identifikaci jednotlivých osob. Tento potřebný soubor si můžeme vytvořit tak, že jako společný klíč použijeme název souboru v původním datasetu. Snímkы v našem datasetu jsem si následně přejmenoval do formátu [identity_id]_[count].jpg.

2.1.2 Výběr snímků

Pro vyhodnocení kvality výsledných deepfakes budu porovnávat deepfaky vzniklé použitím nekvalitního vstupu vůči použití kvalitního vstupu. Nejprve jsme vyřadili snímky s osobami, které v datasetu mají pouze jeden snímek. Poté jsme vybírali snímky odpovídající specifickým charakteristikám – s brýlemi, s překrytím vlasy či jiným překrytím. Tuto činnost bylo nutné provést ručně, protože ji nelze automatizovat. Charakteristiky pro porovnání byly vybrány tak, aby odrážely běžné výzvy při rozpoznávání obličejů, jako jsou obličeje s brýlemi nebo zakryté vlasy. Ukázka vybraných snímků pro každou charakteristiku je zobrazena v obrázku 2.1.



Obrázek 2.1: Vybrané zdrojové obličeje. V první řadě snímky s brýlemi. V druhé řadě snímky s překrytím vlasy. V poslední řadě snímky s překrytím jiného druhu.

V případě deepfaků nedává smysl provádět deepfake tváře ženy na mužskou hlavu a opačně, proto bylo zapotřebí rozdělit použitý dataset na obličeje mužů a žen. Pro dosažení alespoň částečně významného statistického vzorku byla velikost finálního datasetu ustavovena na 30 identit žen a 30 identit mužů. Vedle těchto vybraných identit, ze kterých se budou dělat deepfaky, pak byly dále z datasetu náhodně vybrány dva cílové snímky (muž a žena), u kterých na místo původních obličejů budou dosazeny naše zdrojové obličeje. Tyto cílové snímky byly vybrány s důrazem na dobrou kvalitu (dobře osvícen obličeje bez jakýkoliv obstrukcí), za účelem minimalizace vlivu na kvalitu výsledků – v projektu chceme testovat vliv výběru zdrojových snímků. Vybrané cílové snímky v obrázku 2.2.

2.1.3 Transformace vybraných snímků

Jelikož v projektu na generování deepfaků používáme nástroj FaceFusion (více o projektu FaceFusion v sekci 2.2.1), je nutné při přípravě dat brát v potaz specifické požadavky daného nástroje.

Při prvotním testování funkčnosti byly problémy s tím, že pro některé zdrojové snímkы nedocházelo k vytvoření deepfaků, přičemž u některých to přímo psalo chybovou hlášku ve stylu „Source file not found.“

Během diskuzí s vývojářem bylo zjištěno, že nástroj má problémy s detekcí obličejů na snímcích, kde jsou obličeje částečně oříznuty kraji snímku. Tento problém lze ale jednoduše obejít tak, že kolem snímku se přidá místo navíc, třeba v podobě bílého okraje.



Obrázek 2.2: Vybrané cílové obličeje

Pro vyřešení tohoto problému byl použit nástroj `imagemagick`, kterým jsme ke každému snímku přidali bílý okraj. Tímto způsobem jsme zvětšili velikost obrázků z původních 1024 x 1024 pixelů na 1500 x 1500 pixelů. Po této transformaci nástroj FaceFusion fungoval podle očekávání.

2.2 Deepfakes

Pro účely tvorby deepfakes byly hledány nástroje volně dostupné s otevřeným zdrojovým kódem na serveru GitHub [1]. Byly testovány nástroje jako `faceswap` [5] a `DeepFaceLab` [3], ale ani jeden z nich nevyhovoval našim potřebám. `Faceswap` se neukázal jako efektivní pro jednotlivé snímky a `DeepFaceLab` již není aktivně vyvíjen a jeho zprovoznění bylo komplikované.

Nástroj `faceswap` se nám podařilo zprovoznit bez větších problémů. Avšak tento nástroj není primárně určen pro deepfaking jednotlivých snímků, ale spíše pro práci s většími datasy nebo videi. Součástí funkcionality je natrénování vlastního modelu, nicméně pro to je nutné mít k dispozici dostatečný velký dataset (vyšší desítky či stovky snímků obličeje identické osoby). Je pravděpodobné, že by tento nástroj dokázal produkovat lepší výsledky při použití na videu. Při snaze o vytvoření deepfaku z jednoho snímku na jiný bylo dosaženo velmi špatných výsledků.

U nástroje `DeepFaceLab` bylo ukončeno pokračování vývoje již v roce 2020. Zprovoznění tohoto nástroje bylo velmi komplikované. Tento nástroj funguje na podobném principu jako `faceswap`, kde součástí je natrénování vlastního modelu. Součástí nástroje se nedistribuuje žádný genericky předtrénovaný model.

Nakonec jsme se rozhodli použít nástroj FaceFusion, protože lépe vyhovoval našim specifickým požadavkům na jednoduchost použití a schopnost zpracovávat jednotlivé snímky.

2.2.1 FaceFusion

FaceFusion [4] je poměrně nový (2023) nástroj pro výměnu tváře (deepfakes) nebo vylepšení snímku tváře.

Informace o nástroji

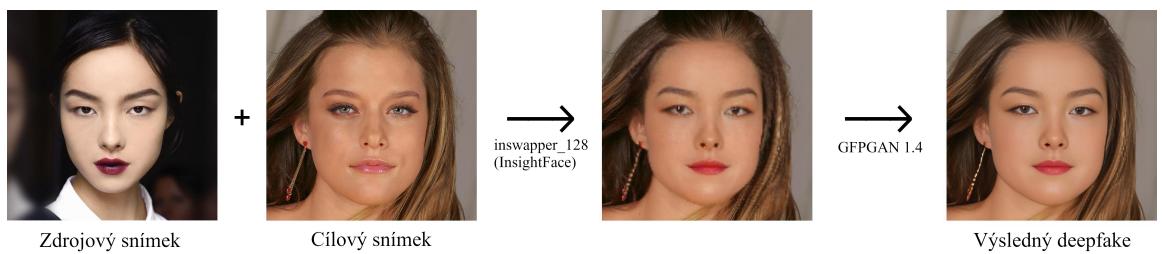
Základem funkcionality nástroje FaceFusion je projekt InsightFace [8], který je považován za nejlepší nástroj na trhu pro zpracování 2D a 3D obličejů. Tento projekt implementuje širokou škálu algoritmů pro rozpoznání obličeje, detekci obličeje a zarovnání obličeje.

FaceFusion používá 128-bitový InsightFace model. Z důvodů etiky projekt InsightFace neposkytuje větší modely jako jsou 256 a 512 bitové

Výsledný deepfake vycházející z použití pouze modelu nedosahuje moc dobré kvality. Nástroj vytváří znatelně rozmazené smínky. Nejedná se ale o neřešitelný problém.

FaceFusion tento nedostatečně kvalitní výstup řeší zpracováním algoritmem pro vylepšení tváře. Nejčastěji k tomu používá algoritmus GFPGAN (nyní ve verzi 1.4). GFPGAN [7] je označení pro praktický algoritmus pro obnovu tváře. Dá se použít pro obnovení černobílých, poškozených a rozmazených fotografií.

Ukázka postupu zpracování a tvorby deepfake 2.3.



Obrázek 2.3: Ukázka postupu tvorby deepfake.

Zprovoznění nástroje

Pro zprovoznění FaceFusion je potřeba mít nainstalovaný Python 3.10 a FFmpeg [6]. Pro grafickou akceleraci je doporučeno mít CUDA Toolkit (>=11.8), ale nástroj může fungovat i pouze na CPU. Instalace potřebných Python knihoven probíhá v lokálním virtuálním prostředí. Instalace potřebných Python knihoven by pak vypadala následovně.

```
python3.10 -m virtualenv .env
. .env/bin/activate
python -m pip install -r requirements.txt
```

Spuštění a výstup nástroje

FaceFusion disponuje přehledným grafickým uživatelským rozhraním, které lze jednoduše spustit příkazem `python run.py`. Nástroj jde samozřejmě spustit i přes příkazovou řádku. Příklad spuštění

```
python run.py -s source.jpg -t target.jpg -o result.jpg --frame-processors \
face_swapper face_enhancer --face-swapper-model inswapper_128 \
--face-enhancer-model gfpgan_1.4 --execution-providers cpu --headless
```

Ukazky různých výstupů je možné si prohlédnout v obrázku 2.4.



Obrázek 2.4: Ukázky výsledných deepfakes. Jednotlivé deepfakes vždy přísluší zdrojového obrázku nad ním.

2.3 Face recognition

2.3.1 CompreFace

CompreFace je open-source nástroj pro detekci, rozpoznání a verifikaci obličejů ze snímku.

Informace o nástroji

CompreFace používá modely FaceNet a InsightFace. Umožňuje detekci obličejů ve snímcích, rozpoznání obličejů ze snímků, které se porovnají s databází a verifikaci dvou obličejů, jestli se jedná o stejnou osobu. Po instalaci a spuštění je několik možností použití. Nejjednodušší možnost je použití webového rozhraní, ve kterém se vše nastaví a otestuje. Nástroj také obsahuje několik SDK pro různé jazyky, kterými lze vytvářet například skripty pro zpracování datasetu, jako v našem případě. Pro určení podobnosti vrací hodnotu similarity, kde se hodnota blíží 1 při porovnání dvou identických obličejů.

Dostupnost a instalace nástroje

CompreFace je dostupný na oficiálním Github repozitáři¹. Oproti jiným nástrojům pro rozpoznávání obličejů nabízí CompreFace jednoduché použití a integraci, což je zvláště výhodné pro výzkumné a vývojové projekty.

Pro instalaci CompreFace stačí stáhnout zip archiv², rozbalit ho a ve složce `s docker-compose.yml` spustit příkaz `docker-compose up -d`. Tento příkaz připraví a spustí Docker kontejner. Pro další spuštění postačí příkaz `docker-compose start`. Pro ukončení běhu kontejneru je možno použít příkaz `docker-compose stop`. Případně je možno použít Docker Desktop (GUI pro Docker) pro manipulaci s kontejnerem.

Po prvním spuštění je třeba CompreFace nastavit ve webovém rozhraní, které je ve výchozím nastavení přístupné na adresě <http://localhost:8000/login>. Při prvním spuštění je nutno vytvořit administrátorský účet. Po vytvoření a následném přihlášení je již možno vytvářet a spravovat aplikace a služby CompreFace. Pro účely tohoto projektu je třeba vytvořit aplikaci se službou typu **VERIFICATION**. Funkčnost lze otestovat rozkliknutím služby a kliknutím na test. Pak stačí nahrát dva soubory, které se porovnají a CompreFace vypíše similarity (podobnost) spolu s dalším údaji, které zjistil.

¹<https://github.com/exadel-inc/CompreFace/tree/master>

²<https://github.com/exadel-inc/CompreFace/releases/tag/v1.2.0>

Zpracování datasetu

Pro vyhodnocení podobností v našem datasetu tvořeném referenčními fotografiemi spolu s částečně zakrytými referenčními fotografiemi a deepfaky, jsme vytvořili skript v jazyce Python. Pro použití CompreFace v Pythonu je dostupné SDK³.

Pro zpracování jsme použili Python verzi 3.10. Na odzkoušení je vhodné mít vytvořeno virtuální prostředí pro Python, do kterého se budou instalovat potřebné balíky. Toto lze provést příkazem `virtualenv .venv`, který vytvoří prostředí do složky `.venv`.

Poté je třeba prostředí aktivovat příkazem `source .venv/bin/activate` (ve Windows `.venv\Scripts\activate`).

Pro instalaci všech potřebných balíků je připraven soubor `requirements.txt` ve složce `CompreFace_scripts`. V této složce se příkazem `pip install -r requirements.txt` nainstalují všechny potřebné balíky.

Spuštění nástroje

Před spuštěním nástroje je třeba v souboru `CompreFace_scripts/compare_deepfakes.py` změnit konstantu `API_KEY` na klíč vytvořené verifikační služby, který je dostupný ve webovém rozhraní (<http://localhost:8000/>). Ve vytvořené aplikaci je vidět seznam služeb s jejich API klíči.

Pro zpracování datasetů ve složce `selection` stačí použít příkaz (z root složky projektu) `python CompreFace_scripts/compare_deepfakes.py selection/men n`.

Skript přijímá dva povinné poziciční argumenty. Prvním argumentem je cesta k datasetu (`selection/men`) a druhým argumentem pak, zda má zpracovávat deepfaky nebo referenční snímek osoby se snímky s různým zakrytím. Pro porovnání deepfaků s referenčními obličeji je možno použít příkaz

```
python CompreFace_scripts/compare_deepfakes.py selection/men y.
```

V obou případech skript navíc porovná všechny referenční obličeje (v souborech s názvem ve tvaru `*_ref.jpg`) mezi sebou, aby bylo možno zjistit, jaké hodnoty nástroj vypočítá, pokud má porovnat dva různé obličeje. Výsledky jsou zapsány ve formě json souborů do každé podsložky datasetu. Pro vytvoření grafu ze získaných dat je k dispozici skript `CompreFace/plot_results.py`. Graf se vytvoří použitím příkazu

```
python CompreFace/plot_results.py selection/men men.png \
    "References CompreFace men"
```

Po dokončení příkazu bude vytvořen graf v souboru `men.png`. Tento skript přijímá tři povinné poziciční argumenty.

1. Cesta k datasetu
2. Název výstupního souboru
3. Titulek grafu (doporučuji napsat mezi uvozovky)

Stejným způsobem lze zpracovat i ženskou část datasetu, v cestě stačí upravit argument `selection/men` na `selection/women`.

³<https://github.com/exadel-inc/compreface-python-sdk>

Výstup nástroje

Výstupem CompreFace jsou data ve formátu json souborů uložená u jednotlivých položek datasetu, ze kterých je možné vytvořit graf ukazující, jak jsou jednotlivé kategorie referenčních obličejů a deepfaků podobné.

2.3.2 MagFace

MagFace je open-source nástroj pro rozpoznání obličejů napsaný v jazyce Python.

Informace o nástroji

Pro tento projekt jsme také využili MagFace s modelem ArcFace pro zjištění podobnosti obličejů. Tento nástroj se používá, jako modul do jazyka Python. Vrací hodnotu euklidovské vzdálenosti pro určení podobnosti, kde hodnota směřující k nule značí vysokou podobnost.

Dostupnost nástroje

Tento nástroj je dostupný na Githubu⁴, my ale využili verze dodané vedoucím projektu.

Instalace a použití nástroje

Tento nástroj potřebuje větší množství Python modulů. Instalace těchto modulů může tedy zabrat dost velké množství času. Pro běh tohoto nástroje je vhodné vytvořit separátní virtuální prostředí, aby nenastaly konflikty mezi instalovanými balíky. Tento nástroj jsme používali s Python verzí 3.10. Po vytvoření a spuštění virtuální prostředí (příkazem `virtualenv .venv` a `source .venv/bin/activate`) se potřebné balíky nainstalují příkazem `pip install -r requirements.txt`.

Po instalaci je možno nástroj použít příkazem

```
python MagFace/compare_dataset.py selection/men n
```

pro provnání referenčních obličejů s jejich lehce zakrytými variantami. Při nastavení posledního argumentu na `y` nebo `yes` se referenční obličeje porovnají s vytvořenými deepfaky. Celý příkaz je:

```
python MagFace/compare_dataset.py selection/men y
```

Po vygenerování těchto výsledků je třeba také spustit příkaz

```
python MagFace/cross_check.py selection/men
```

Tento příkaz porovná referenční obličeje mezi sebou, aby bylo jasné, jaké hodnoty budou u dvou různých obličejů. Posledním krokem je vytvoření grafů příkazem:

```
python MagFace/plot_results.py selection/men magface_ref_men.png \
    "MagFace references men"
```

Pro provnání ženské části datasetu stačí místo argumentu `selection/men` poslat argument `selection/women`.

Výstup nástroje

Výsledkem porovnání jsou data ve formě json souborů v jednotlivých složkách datasetu. Tato data jsou dále zpracována do grafů.

⁴<https://github.com/IrvingMeng/MagFace>

Kapitola 3

Experimentování a vyhodnocení

Předmětem této kapitoly bude tvorba deepfakes a následná evaluace jejich kvality. Budeme zkoumat různé aspekty kvality deepfaků, včetně jejich realističnosti a schopnosti přesvědčit, a použijeme k tomu nástroje pro rozpoznání obličeje. Metody a postupy použité při tvorbě deepfaků a jejich hodnocení budou podrobně popsány.

3.1 Tvorba deepfakes

Dataset ve své finální podobě sestává z vybraných snímků mužů a žen, které byly zvětšeny na velikost 1500 x 1500 pixelů přidáním bílého okraje. Před spuštěním bylo nutné zvolit správné hodnoty argumentů nástroje FaceFusion. Výběr modelu *inswapper_128* a *GFGAN_1.4* byl motivován snahou dosáhnout co nejlepší kvality výsledků.

Nástroj FaceFusion pak spustíme následovným způsobem:

```
python run.py -s source.jpg -t target.jpg -o result.jpg --frame-processors \
face_swapper face_enhancer --face-swapper-model inswapper_128 \
--face-enhancer-model gfgan_1.4 --execution-providers cpu --headless
```

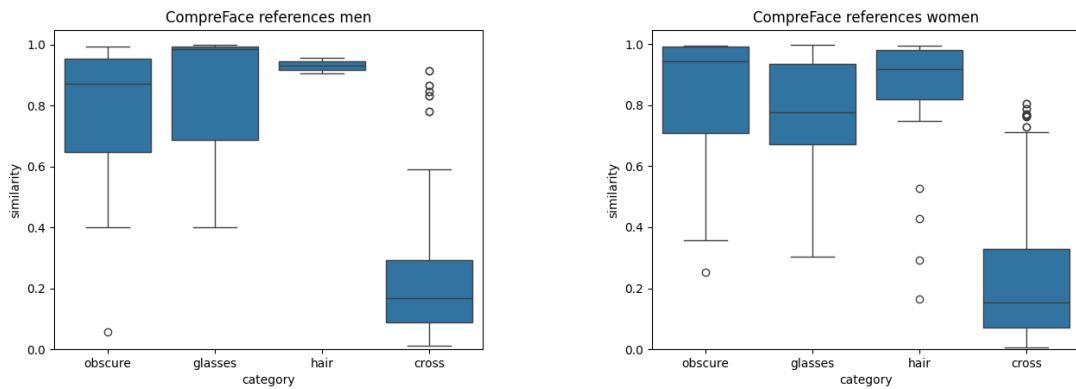
Ukázka výsledných výstupu je v obrázku 2.4. Po vytvoření deepfaků následuje jejich evaluace, kde budeme zkoumat různé aspekty kvality, jako je realističnost a schopnost přesvědčit.

3.2 Rozpoznání obličejů prvním nástrojem

Primární výstupní hodnotou porovnání nástrojem CompreFace je míra podobnosti (similarity) udávaná v procentech (jako desetinné číslo). Teoreticky tak míra podobnosti 1 by měla označovat identickou osobu nade vši pochyby. S klesající hodnotou k nule pak zase obličeje, které se značně liší.

V našem datasetu jsme pro každou osobu definovali referenční snímek a pomocí CompreFace jsme porovnali podobnost mezi různými typy snímků a tímto referenčním. Výsledky jsme vizualizovali v krabicových grafech, kde každý sloupec reprezentuje určitý typ snímku (např. s brýlemi, s vlasy, atd.). Dostaneme tak například výstup zobrazený na obrázku 3.1. Všechny sloupce v obrázku 3.1 jsou vysvětleny následovně:

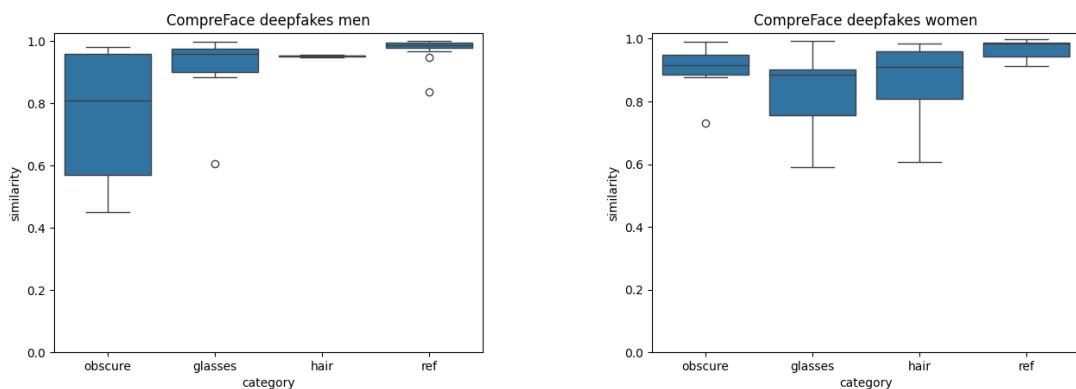
- obscure – Porovnání reference se snímkem s různě zakrytými částmi obličeje
- glasses – Porovnání reference se snímkem osoby s brýlemi



Obrázek 3.1: Míry podobnosti mezi vstupními snímky mužů a žen.

- hair – Porovnání reference se snímkem osoby s obličejem zakrytým vlasy
- cross – Porovnání referenčních snímků mezi sebou (zde je očekávána hodnota blízká 0)

Jak je možno vidět, obličeje dvou různých osob, se s hodnotou similarity pohybují mezi 0,1 a 0,4, zatímco stejné obličeje, které jsou nějak zakryté mají rozmezí spíše 0,7 až 0,99. V případě porovnání dvou různých obličejů se v datasetu nachází několik outlierů. Výsledky naznačují, že některé typy zakrytí mají větší dopad na schopnost nástroje správně identifikovat obličeje. Například snímky zakryté vlasy dosahly vyšších hodnot podobnosti než snímky s brýlemi. Nástrojem CompreFace se dále vyhodnotí podobnosti referenčního snímku s deepfaky, které byly vytvořeny z snímků, které jsou nějak zakryty (slunečními brýlemi, vlasy, mikrofonem před ústy, atd.). Výsledky je možno nalézt v grafu 3.2.



Obrázek 3.2: Míry podobnosti mezi vytvořenými deepfaky a referenčními snímky mužů a žen

Sloupce v obrázku 3.3 nyní zachycují podobnost mezi deepfaky vzniklé z příslušných zdrojových snímků s referenčním snímkem (originálem). Nový sloupec *ref* zaznamenává porovnání mezi výslednými deepfaky, které vznikly z referenčních snímků, tedy ze snímků se kterými se porovnává. Jedná se o kontrolní metriku.

3.3 Vyhodnocení dat

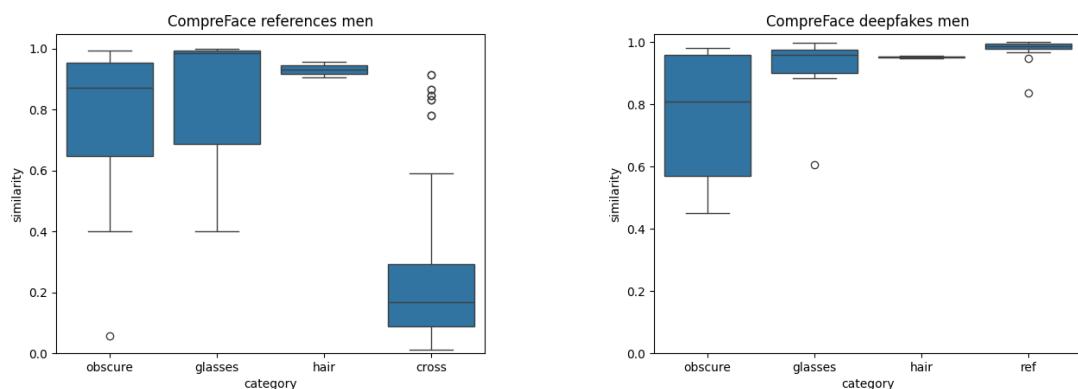
Všechny statistické data všech naměřených dat byly zaznamenány do tabulky 3.1. Při vyhodnocení dat se bude ale primárně vycházet z krabicových grafů. Než je možné hodnotit

zdroj										deepfake							
obscure		glasses		hair		cross		obscure		glasses		hair		ref			
med	std	med	std	med	std	med	std	med	std	med	std	med	std	med	std	med	std
ženy muži		0.87	0.3	0.99	0.25	0.93	0.04	0.17	0.18	0.81	0.21	0.96	0.12	0.95	0.01	0.99	0.04
		0.94	0.30	0.78	0.24	0.92	0.23	0.15	0.20	0.92	0.08	0.88	0.13	0.91	0.13	0.19	0.23

Tabulka 3.1: Výsledná statistická data všech naměřených výsledků z nástroje CompreFace

kvalitu výstupních deepfaků je nutné nejdříve porovnat „kvalitu“ vstupních snímku. Lze tak stanovit referenční obor hodnot výstupu u nástroje pro rozpoznání obličeje a zároveň přehled o vhodnosti vstupních snímků. V obrázku 3.3 sloupec *cross* zachycuje porovnání mezi referenčními snímky všech osob navzájem. Většina hodnot se pohybuje kolem míry 0.2. To nám značí, že tento nástroj dokáže produkovat hodnoty na celé škále od 0 do 1. Pár outlierů se pohybuje kolem hodnoty 0.8. To je ale očekávané chování, neboť některé osoby jsou si více podobny.

Sloupec *ref* u grafu sledující deepfaky zobrazuje hodnoty soustředěné velmi blízko jedničce. Jedná se o kontrolní metriku, která by měla být velmi blízká jedničce, takže je vše v pořádku.



Obrázek 3.3: Míry podobnosti zdrojových snímků (vlevo) a výsledných deepfaků (vpravo) s referenčním snímkem pro muže.

Kategorie *hair* se u zdrojových snímků se nacházela kolem mediánu 0,93. Při zhotovení deepfaků ze snímků obličejů částečně zakrytých vlasy se medián posunul na 0,95. Došlo tedy ke zlepšení.

Kategorie *glasses* u zdrojových snímků měla velmi vysoký rozptyl i mezikvantilový rozsah hodnot. Evidentně se tak jednalo o problematický vstup pro nástroj CompreFace. Pozoruhodný výsledek tak dostáváme při vytvoření deepfaků nástrojem FaceFusion. Při zhod-

nocení míry podobnosti u takových deepfaků je evidentní, že mezikvantilový rozdíl i rozptyl se zásadně zlepšili.

Kategorie *obscure* představuje očekávaný výsledek. Medián klesl z 0,87 na 0,81. Míra variability dat zůstává podobná. Z toho vyplývá, že některé typy a míra překrytí části obličeje dokáže být zásadním problém při tvorbě deepfaků, a nejedná se o vhodné zdrojové snímky.

Nicméně ve zdrojových snímcích pro kategorii *obscure* se nacházel jeden outlier. Při zpracování nástrojem FaceFusion se tento outlier přesunul do standardního rozptylu. Ná-

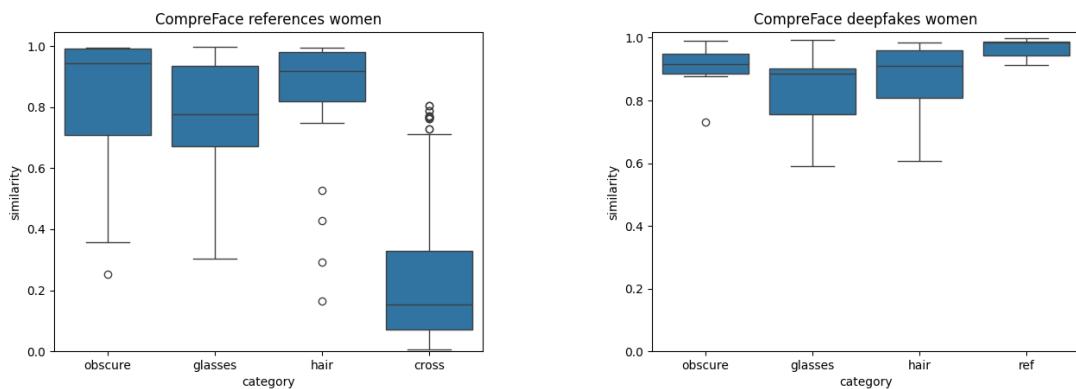


Obrázek 3.4: Ukázka outlieru ve zdrojových snímcích a jeho výsledný deepfake.

stroj FaceFusion provedl i z tak nevhodného snímku deepfake, který by se možná dal označit za použitelný. Obdivuhodný výsledek.

Výsledky porovnání snímku žen

Výsledky výše platí pro snímky mužů. Snímky mužů a žen se zásadně liší. Proto je nutné se podívat i na grafy zachycující snímky žen v obrázku 3.5.



Obrázek 3.5: Míry podobnosti zdrojových snímků (vlevo) a výsledných deepfaků (vpravo) s referenčním snímkem pro ženy.

Graf míry podobnosti vstupních snímků (vlevo) v obrázku 3.5 vykazuje podobné hodnoty jako v případě mužů. Největším rozdílem je pravděpodobně sloupec *hair*, který vykazuje větší rozptyl hodnot a několik outlierů. Tento rozdíl je jednoduše vysvětlitelný faktem, že ženy pravděpodobně mívají delší vlasy, které tak budou zakrývat větší část obličeje.

Ukázka všech čtyřech outlierů ve sloupci *hair* je vidět v obrázku 3.6. Porovnáme-li ale



Obrázek 3.6: Ukázka outlierů u zdrojových snímků s vlasy v případě žen.

v obrázku 3.5 hodnoty pro výsledné deepfaky žen s předešlými výsledky deepfaků pro muže (obrázek 3.3) dostáváme nižší míru zlepšení výsledků. Medián pro *hair* se prakticky nezměnil, u mužů klesnul o 2 %. Medián sloupce *glasses* u žen si polepšil o 13%.

Dostáváme ale částečně překvapivý výsledek u kategorie *obscure*. Rozptyl u této kategorie klesnul z 0,3 na 0,08.

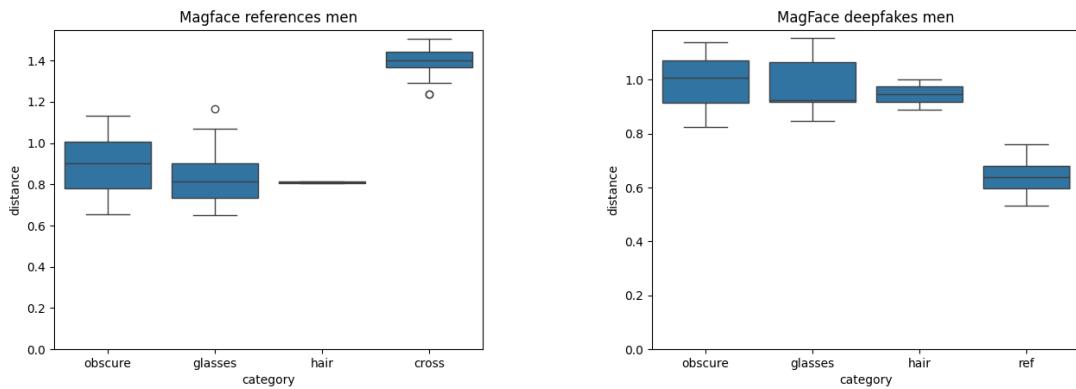
Závěr

Výsledky CompreFace ukázaly, že v případě mužů dosahovaly snímky s vlasy a brýlemi vyšší míry podobnosti než u zakrytých obličejů. Zajímavě, i v případě žen byly zaznamenány podobné trendy, přičemž výsledky ukazují na větší variabilitu u snímků s vlasy. Tyto výsledky mohou naznačovat, že nástroj FaceFusion je velice efektivní ve tvorbě deepfaků, nebo nemáme nedostatečně dobrý nástroj pro rozpoznání obličeje. Proto by bylo ideální provést stejnou kontrolu, ale za použití jiného nástroje pro rozpoznání obličeje.

3.4 Zkvalitnění dat použitím dalšího nástroje

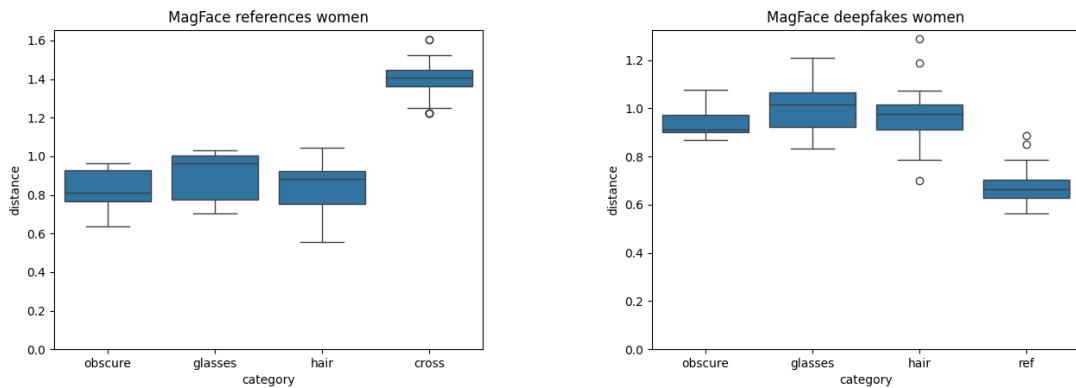
Výstupem nástroje MagFace není míra podobnosti (similarity) jako u CompreFace, ale euklidovská vzdálenost vektorů. Mění se nám tak obor hodnot a směr grafu. Čím nižší vzdálenost tím větší podobnost obličejů, čím vyšší vzdálenost tím méně jsou si porovnávané obličeje podobny. V obrázku 3.8 se hodnoty sloupce *cross* pohybují kolem mediánu 1,4. Tedy vzdálenost pro odlišné obličeje se blíží k hodnotám 1,4. Kontrolní sloupec *ref* zase zobrazuje hodnoty blížící se k 0,5. Referenční hodnoty při vyhodnocení deepfaků by měly ukazovat velkou podobnost. Výstupem nástroje MagFace tedy pravděpodobně budou hodnoty v rozmezí (0,5;1,5).

MagFace v případě deepfaků mužů naznačuje snížení podobnosti ve všech kategoriích. Nicméně hodnoty se ustalují kolem 1,0, tedy stále jsou si obličeje na snímcích podobny, ale nástroj jim nepřiřazuje optimistickou pozitivitu.



Obrázek 3.7: Výstupní hodnoty nástroje MagFace pro zdrojové snímky a výsledné deepfaky mužů.

Výsledky z nástroje MagFace pro ženy vykazují podobný trend. Vytvořením deepfaků klesla podobnost obličejů, ale neobjevuje se žádná preference pro určitý typ kategorie vstupních snímků.



Obrázek 3.8: Výstupní hodnoty nástroje MagFace pro zdrojové snímky a výsledné deepfaky žen.

3.5 Vyhodnocení výstupu

Výsledky z nástroje MagFace pro muže ukazují, že i když podobnost obličejů klesla ve všech kategoriích, hodnoty se stále pohybují kolem 1.0, naznačující určitou míru podobnosti. Pro ženy tyto výsledky naznačují podobný trend, ale bez zjevné preference pro jakýkoliv typ vstupního snímku.

Rozdíl výsledků mezi nástroji CompreFace a MagFace by mohl být vysvětlitelný. CompreFace obecně byl více nakloněn věrohodnosti deepfaků než MagFace. Takovéto výsledky by mohly vycházet z toho, že jak FaceFusion, tak CompreFace využívají pro rozpoznání obličejů model InsightFace, tedy oba nástroje hledají v daném snímku stejné rysy.

Kapitola 4

Závěr

V našem projektu jsme se zaměřili na generování deepfakes s využitím nejnovějších nástrojů (konkrétně FaceFusion), které využívají moderní modely pro rozpoznání obličeje jako je InsightFace a algoritmy pro zlepšení snímku tváře jako je GFPGAN.

Naše zjištění naznačují, že i s nekvalitními zdrojovými snímky lze vytvářet přesvědčivé deepfakes, zároveň ale bylo zjištěno že určité prvky omezující viditelnost částí obličeje mají zásadní vliv na kvalitu konečného výstupu.

Tato práce má ale několik omezení. Prvním je kvalita a rozsah datových sad. V této práci se používá datová sada CelebA-HQ. Tato datová sada ale obsahuje pouze snímky o rozlišení 1024 x 1024 pixelů. V reálném světě je každý snímek unikátní a různé snímky o různém rozlišení a o různé kvalitě (vhodnosti) dokáží produkovat značně odlišné deepfakes o různé kvalitě.

Další omezením je volba použitých nástrojů pro rozpoznání obličejů. Různé nástroje pro rozpoznání obličeje generují odlišné výsledky především kvůli variacím v použitých algoritmech a datových sadách. Například nástroje založené na hlubokém učení mohou identifikovat a analyzovat obličejové rysy s větší přesností než ty, které používají tradiční metody strojového učení.

Pro další výzkum by bylo vhodné použít širší škálu datových sad a porovnat výsledky s použitím různých nástrojů pro rozpoznání obličejů, aby se získal ucelenější obraz o schopnostech a omezeních současných technologií deepfake.

Zjištění v této práci mohou mít dopady na oblast bezpečnosti a autenticity digitálního obsahu. Ukazují, že je možné vytvořit přesvědčivé deepfakes i s nekvalitním zdrojovým snímkem, což má důležité implikace pro oblasti jako jsou média, zábava a bezpečnostní služby. Naše zjištění tak může být varováním pro uživatele sociálních sítí a digitálních platform ohledně možných rizik spojených s deepfakes.

Moderní nástroje pro tvorbu deepfakes dosahují obdivuhodných výsledků. I v případech, kdy má nástroj k dispozici zdrojový snímek označitelný za ne moc vhodný, jsme schopni dosáhnout velmi dobrého výsledku.

Literatura

- [1] *GitHub* [online]. [cit. 2023-11-17]. Dostupné z: <https://github.com>.
- [2] *CelebAMask-HQ* [online]. 2020 [cit. 2023-11-17]. Dostupné z:
<https://github.com/switchablenorms/CelebAMask-HQ>.
- [3] *DeepFaceLab* [online]. 2020 [cit. 2023-11-17]. Dostupné z:
<https://github.com/iperov/DeepFaceLab>.
- [4] *FaceFusion* [online]. 2023 [cit. 2023-11-17]. Dostupné z:
<https://github.com/facefusion/facefusion>.
- [5] *Faceswap* [online]. 2023 [cit. 2023-11-17]. Dostupné z:
<https://github.com/deepfakes/faceswap>.
- [6] *FFmpeg* [online]. 2023 [cit. 2023-11-17]. Dostupné z: <https://www.ffmpeg.org/>.
- [7] *GFPGAN* [online]. 2023 [cit. 2023-11-17]. Dostupné z:
<https://github.com/LexKoin/GFPGAN-1.4>.
- [8] *InsightFace* [online]. 2023 [cit. 2023-11-17]. Dostupné z:
<https://github.com/deepinsight/insightface>.